

Guide to ISI Applets

(<http://www.isi-stats.com/isi2nd/ISlapplets2021.html>)

Notes: Many of the applets now come with Spanish translations. You can follow a link at the bottom of the applet. Let us know if you need others or want to suggest revisions to the translations. Versions of the applets that will load in their own windows can be found at <http://www.isi-stats.com/isi2nd/applets2021.html>. Additional applets can be found at www.rossmanchance.com/applets2021/.

One Proportion

The One Proportion applet focuses on inference (testing) for a single categorical (binary) variable. The applet default is to focus on probability of heads and number of tosses, and the animation displays coin tossing. If the user changes the probability of heads to a number other than 0.5, then the applet switches to probability of success and number of samples, and the animation displays spinners. (This [version](#) uses generic language the whole time.) The **Animate** box can be unselected if you want to focus on the outcomes of individual trials (e.g., keep number of repetitions set to one and press the **Draw Samples** button multiple times to visually see the sample to sample variability). Animation will also turn off automatically if you set the **Number of repetitions** to a value larger than one. The results of the last repetition are then sorted into successes (or heads) and failures (or tails) below the horizontal line for a type of bar graph of the simulated results. A “Total” appears under the Draw Samples button displaying the number of repetitions/samples so far. You can check the “Hide Coins/Hide Spinners” box on the far right to not show this output (and move up the dotplot for more compact output).

There is also a check box (**Summary Stats**) to display the mean and standard deviation of the generated distribution in the upper left corner of the dotplot. The statistic can be either the number of heads/successes or the proportion. You will need to match this format in the **As extreme as** box or the applet will give you a warning. The **As extreme as** proportion does not need to be a possible outcome (e.g., for informal continuity corrections) so you may want to be careful with rounding. You can also press the direction toggle to switch from $\geq$ to $\leq$.

By checking the **Two-sided box** the displayed Count Samples output will sum the probabilities of all outcomes with a smaller probability than the one observed based on the exact binomial probabilities (in which case the applet does round any non-integer inputs to the most appropriate values depending on the direction toggle). After checking this box, you also have the option of checking the “Between” box and specifying a second value (both endpoints are inclusive). By checking the **Exact Binomial** box this probability will be displayed along with green lines displaying the pmf on the dotplot and the output in terms of X values. You can check the **Normal Approximation** box to display this probability (assuming the theoretical mean and SD values) along with overlaying the pdf on the dotplot and the output in terms of Z values. (These last two probabilities can be obtained even without running a simulation first.)

The vertical line that is displayed in the dotplot corresponding to the inputted value for the observed statistic is also moveable, updating the p-value approximations automatically.

Note: If you have a touch screen computer, often the touch screen will work to move items but not the mouse.

The **Show previous** box will display a “ghost distribution” for the previous settings of the input parameters (e.g., changing sample size). The **Show sliders** box allows you to slide values for the process probability and the sample size, updating the picture automatically. If you have inputted a value in the “As extreme as” box, the axis will stay anchored to that value.

Another statistic, the longest run on successes/heads can also be selected at the bottom of the applet.

A link to a Spanish version or a German version of the applet can be found at the bottom. You can also go to versions that model sampling Reese’s Pieces candies or colored candies.

Two Proportions (see also Multiple Proportions)

Using [this URL](#) defaults to the user inputted cell counts for a generic two-way table (otherwise you may need to check the **2x2** box). (Pre-filled in data matches the dolphin therapy study, Ch. 5. Other options for default data are listed at the bottom.)

- If typing in counts into the 2x2 box, press **Use Table**.
- If pasting in two columns of data, press **Use Data**.
- If typing in a two-way table (use one-word row and column labels and no totals), press **Use Table**, see example below.

You can view the resulting table, and the conditional proportions, by checking the **Show Table** box. The segmented bar graph will display and you can select which outcome is treated as success. You can change the order of subtraction (i.e., which is Group A) by pressing the **(GroupA-GroupB)** button. You can also use the pull-down menu to switch to a mosaic plot.

(If you uncheck the 2x2 box, you can change the EV and RV names to match your study.)

The choice of statistic can be changed using the **Statistic** pull-down menu on the left. (Summer 2020: MAD has been replaced with Mean Group Diff.) You can also display the chi-square table with chi square cell contributions and p-value, as well as two sample z confidence intervals. (There are also options for “cell 1 count”, relative risk, and odds ratio (when have 2x2 data) and associated confidence intervals, as well as displaying the log transformed statistics. There is also an option for the exact p-value from Fisher’s Exact Test.)

To perform a randomization test, check the **Show Shuffle Options** box. If you are using the 2x2 box, you have the option of displaying blue and green “index cards” to represent the outcomes.

Pressing **Shuffle** will visually mix up the cards and deal them out to the two explanatory variable groups. The dark line at the top of the cards represents cards that originally began in “Group A.” You can also display either the raw data (showing the randomized RV values) or the segmented bar graph of the simulated data. The output for the most recent shuffle will display below (using blue headings).

The applet also automatically displays the mean and standard deviation of the simulated statistics. The user needs to specify the value to count beyond (use the toggle button to change the direction from Greater Than (or equal to) to Less Than (or equal to) or Beyond (for two-sided p-values). The applet will complain if the inputted value does not match the observed value of the statistic but will proceed to display the output under the graph. Click on a “box” in the histogram will display the corresponding simulated table.

There is a **Show Previous** box but it works best if you make small changes to the two-way table. The link for the [yawning](#) data is slightly modified data, this [link](#) is for the actual data.

Multiple Proportions

This applet (<http://www.rossmanchance.com/applets/ChisqShuffle.htm>) analyzes two or more proportions (a 2x2 table or larger two-way table). The default data is from the Organ Donor study (Ch. 8). The user can **Clear** the sample data and paste in either raw data (and then press **Use Data**) or a two-way table, e.g., of the form

```
dolphin    control
improve    10    3
not 5      12
```

(using spaces between the columns if typing in or can use tabs from pasted data) and then pressing **Use Table**. Use single words for variable names. The results of the table are shown if the **Show table** box is checked but you may want to edit the table to improve formatting (e.g., shorten column names, align columns, delete totals). If you delete the totals, you can then paste this table back in to the box to improve the variable and category names. You can also resize the table by clicking on the lower right corner and dragging in/out. This table will also display the conditional proportions (depending on which outcome you define as success which can be changed with the pull-down menu below the sample graph).

If the applet detects that you are using a 2x2 table the choice of statistic options available change to include difference in proportions and relative risk. (MAD has been renamed Mean Group Diff.)

For the simulation, check the **Show Shuffle Options** box and specify the **Number of Shuffles**. You can choose to display either the re-randomized data values (Data) or the corresponding segmented bar graph (Plot). In this panel of the applet, titles are blue to convey they correspond to the re-randomized data. “Bricks” are added to the dotplot, with the most recently placed brick

in blue, until you have over 100 shuffles when the graph switches to a histogram. The mean and standard deviation of the shuffled statistics are also displayed in the upper right corner of the graph. You can also click on a brick to turn it blue and to display the corresponding shuffled results for that shuffle (you may need to scroll to the top of screen first so the applet recognizes your mouse location). To reset the simulation, uncheck and recheck the Show Shuffle Options box.

The observed value of the statistic (shown in panel one) should be entered into the **Count Samples** box or the applet will complain (but will still do the counting) if a different value (for the observed statistic) is entered. The user can also specify counting values to the right, to the left, or in both tails (more extreme in either direction as determined by outcomes with smaller probabilities than the one observed). Warning messages display if the user doesn't input the observed statistic from the left panel or tries to specify "beyond" with the chi-square statistic.

Depending on the choice of statistic, you can **overlay** a normal or a chi-square pdf and the theoretical p-value result will also be shown. You can also proceed directly to displaying the chi-square output (**Show X^2 output**) which includes the cell contributions to the chi-square statistic and the degrees of freedom.

With two proportions as the statistic and/or when displaying the chi-square output, you can also choose to display confidence intervals for differences in proportions. These match the individual two-sample z-intervals. An asterisk is displayed if the interval does not include zero.

Descriptive Statistics

You can paste in your own data or generate a random sample. If you want to paste in a categorical grouping variable as well, select the **Stacked** box before pasting (a button will appear to specify whether the categorical variable will be the first or second column). If you have variable name(s), make sure there are no spaces between words in a variable name and that you check the **Includes header** box in appropriate. Once you have data, you can first guess the values of variable's statistics by checking the box and moving the line(s). Checking the **Show Percentages** box will display what percentage of the sample data fall within one (guessed) standard deviation of the mean. (With a touch screen, you may need to use the touch screen rather than the mouse to move the lines.)

Note the dotplot often does some binning. You can also change the display to a histogram, with choice of bin width by using the Number of bins slider, and you can also add boxplots. There is a corner tab that allows you to resize the graph.

Checking the **Show Actual** boxes will overlay those values. The Guessed SD lines can be moved to match the actual for the Show Percentages. The quartiles can be found by viewing the calculation for the IQR.

Clicking on a dot will display its value on the Delete button and the observation can be removed and statistics will automatically update. Using the mouse to click on and move a value will automatically update any displayed actual statistics. The **Random sample** button generates data from different population shapes to help illustrate the relationship between and median, impact of outliers etc. You can display instructions for an activity using the **Show instructions** box.

If the observations are ordered, you can toggle to a **Timeplot** using the checkbox. You can also switch to a normal probability plot (observations vs. z-scores; normally distributed data should fall along a line).

One Mean

This applet (aka Sampling from Finite Population) allows the user to copy and paste in multiple columns for a large population of numerical values or of words and then take samples from that finite population. There are three built in populations as well (the context is sleep hours, the populations are symmetric, skewed right, and uniform), as well as radio buttons for the Gettysburg Address, Ages of Pennies (left skewed) and change (uniform), Stars in the sky, and lyrics to a Beyonce song. If you want to view the values in the population window you can use the bottom to return to the **top** or the **bottom** of the window (e.g., for iPads where scrolling is more difficult). After pasting in the population, press **Use Data** to display a histogram and summary statistics for the population (if you paste in words, it displays the distributions of the word lengths). With multiple columns, you need to use the **Variable** pull-down menu to select which variable you want to sample from.

Then check **Show Sampling Options** and specify the **Number of Samples** and **Sample Size**. The selected observations are highlighted in blue in the population histogram. The dotplot in the middle panel shows the distribution of the sample and the dotplot in the rightmost panel displays the sample statistic (choosing between mean, median, and t -statistic using the **Statistic** radio button across the top, or proportion with a categorical variable). The sampling distribution will show bricks (the most recent in blue) until there are over 100 samples and then it converts to a histogram. Clicking on one of the bricks will display the corresponding sample in the Sample graph. You can use the **Population scale** radio button to change the scaling to that of the population. If you chose the **Fixed** scale (newer version), it won't change as you change input options (e.g., sample size); otherwise, use the mean and standard deviation summary statistics (upper left corner) to compare to the population or previous distributions. You can also overlay a theoretical (CLT) normal distribution on the distribution of means or medians or proportions or the corresponding t distribution on the t statistics (allows you to overlay the normal distribution on t statistics as well).

Use the **Count Samples** box to count the observations at least, at most, or at least as extreme (in either tail, symmetrically from the mean of the sampling distribution) from the inputted value

(and the greater than, less than, and beyond pull-down menu). If overlaying a normal or t distribution, the applet will display the theoretical probability as well.

Under Show Sampling Options, if your putted data contains multiple columns you can also specify cluster or stratified sampling. Stratified sampling assumes proportional sampling (e.g., if the categorical variable you specify is 20/80 split, then the resulting sample will have a 20/80 split), color coded in the Sample plot by the categories. For cluster sampling, specify the number of clusters.

The **Population size** radio buttons allow the user to make additional copies of the population so that the population size is 4 times larger or 40 times larger. A slider below the population graph allows the user to shift the center of the distribution left and right before sampling.

Links at the top of the applet allow you to use [bootstrapping](#) (resampling from the sample) or to sample from [theoretical](#) population distributions (normal, skewed right, and uniform) after specifying population parameters (e.g., mean and SD).

Multiple Means

This applet (aka Comparing Groups on a Quantitative Response) performs the randomization simulation for two or more groups. The default data is the Sleep Deprivation study (Ch. 6). You can press **Clear** and then paste in new data (use the button above the data window to toggle the ordering of the columns and the check box to specify stacked vs. unstacked data) and then press **Use Data**. Use single words for variable names. If you paste in more than 2 columns, use the pull-down menus underneath to identify the response and explanatory variables.

The applet displays dotplots and summary statistics (the displayed summary statistics change if you use the difference in medians as the statistic), with the option of showing boxplots as well. The first three options for **Statistic** are not available if you are using more than two groups (and the default statistic becomes the Mean Group Diff, the average difference between all the group mean pairs). Other statistics you can choose from include Max-Min, R^2 , Effect size (dividing by the within group variation), and p-value (ask students to predict the behavior of the distribution first).

The **Show Shuffle Options** works very similar to the Multiple Proportions applet as described above. With two groups, selecting the t -statistic allows you to overlay a t distribution on the randomization graph. Selecting the F -statistic allows you overlay the corresponding F distribution. Warning messages display if the user does not input the observed statistic from the left panel. Pressing on a dot in the randomization distribution will show the corresponding randomization output (you may need to scroll to the top of screen first so the applet recognizes your mouse location). You can specify a non-zero value for the hypothesized difference in the

population mean. The animation (Number of shuffles = 1) shows the treatment effect being subtracted off, then the shuffling, then added back on.

On the left, you can also check the box to **Show ANOVA table** and the box for **95% CIs for difference in means** to display two-sample t -intervals. This displays all pairwise intervals, using the pooled standard deviation, and denotes intervals not containing zero with an asterisk.

Matched Pairs

The default data is from the baserunning study (Ch. 7). You can also paste in two columns of paired data (with or without an initial column of IDs). The applet will display separate dotplots (for list one and for list two) but will connect the outcomes in each pair with a line. A dotplot of the differences will also display, along with summary statistics. If on an iPad, use the **Top/Bottom** button scroll in the data window.

Checking the **Randomize** box and then pressing the **Randomize** button will visually display some of the pairs swapping the order of the values (e.g., if a coin flip comes up heads) and the other pairs not swapping order (e.g., coin flip comes up tails). Then the new positions and differences are displayed in the top two dotplots, and the mean difference is added to the bottom dotplot. (The animation will not fully display once you increase the number of randomizations.)

Use the pull-down menu to **Count Samples** at least, at most, or at least as extreme (in either tail, symmetrically from the mean of the sampling distribution) from the inputted value. If you select the t -statistic (above the dotplot), you can calculate the t -statistic by hand or it will now display in the dotplot of the differences. You also then have the options to overlay the t distribution. Once that is checked, a check box appears on the left to calculate a **95% CI for the mean difference**.

Corr/Regression

This applet (aka Analyzing Two Quantitative Variables) has many options, and you may only want to show a few of them at a time. The default data is for student foot length measurements (cm) and heights (in). You can press **Clear** and then paste in your own data (using one-word variable names), toggle the **Explanatory, Response** button if necessary, and press **Use Data**. If you paste in more than 2 columns of quantitative data, use the pull-down menus underneath to select the response and explanatory variables.

Checking the **Show Movable Line** box adds a blue horizontal line to the graph at the height of the mean of the response variable. Using the mouse to click on the small green boxes at the end of the line and dragging will allow you to change the slope of the line; using the mouse to

click on the larger green box in the middle of the line and moving up and down will allow you to change the intercept of the line. This mouse functionality does not currently work if you have a touch screen, but you should be able to accomplish this using the touch screen. The displayed line equation will automatically update as you move the line.

Checking the **Show Residuals** box draws in the vertical distances from the y-values and the line and displays $SAE = \text{sum of absolute errors (observed} - \text{predicted)}$. Checking the **Show Squared Residuals** box displays squares around these vertical lines and displays the $SSE = \text{sum of squared errors}$. These values will update automatically if you move the line.

You can add and remove observations from the data set. To add, enter the x, y coordinates and press **Add**. To remove, click on the dot – it will turn red and the coordinates will appear on the **Delete** button which you should then press. If you need to return to the original data set, press the **Revert** button below the data window.

The **Show Regression Line** box will then add the least squares line (red) and display its equation. The Show (Squared) Residuals boxes work similarly as for the movable line. If you have not checked the Show Movable Line box, you can also move an individual observation by first checking the **Move observations** box and then clicking on the point and dragging. All regression output will automatically update.

You can also display the **correlation coefficient** and the **R-squared** values for your data, as well as the full **Regression Table**. Once you display the Regression Table, you will also see an option to display the 95% confidence interval for the slope. There is also a checkbox to display the **Regression SE** or the square root of the sum of the square residuals (aka s). You can also view the ANOVA table and the Spearman correlation coefficient (the correlation coefficient of the ranks of each variable).

For **Show Shuffle Options**, you need to specify the **Number of Shuffles** and choose the pull-down menu above the graph for the choice of statistic (will default to slope otherwise). Pressing the **Shuffle Y-values** will mix up the order of the response variable values and reassign them back to the explanatory variable values. If you display “Data” you will see how the order of the y-values has changed but not the x-values. If you display “Plot” you will see the shuffled data and if the slope statistic is selected you will also see the original regression line (in red) and the regression line for the shuffled data (in blue). These lines will accumulate as you carry out more shuffles.

For choice of statistic, you have correlation coefficient, slope, and t -statistic (for slope), and the regression intercept. (You can also view the R -squared statistic and the F -statistic from the ANOVA table.) Using the t -statistic allows you to overlay the theoretical t distribution (and p-value). The user has to decide what value to input to count above (greater than), below (less than), or beyond (symmetric) but gives a warning message if the value does not match the observed statistic. (You can use the Regression Table output to calculate the observed

t -statistic, especially as the standard error used in the Regression Table may not match the standard deviation of the simulated slopes, particularly for stronger associations.)

To perform random sampling rather than random shuffling, check the **Design Population** check box in the lower left. You can specify parameter values (the default values match the sample statistics, including s) and then press the **Create Population** button. You may need to press the **Rescale** button to improve the display (but may not want to later to help students distinguish the changes in the population). In the graph of the sampled regression lines (using the slope as the statistic), the most recent link is dark blue and the specified population regression line is in light blue. A checkbox below allows you to add the original sample's regression line. (Error messages are not provided if the Count Sample input does not match this value.) In the **Design Population** pull-down menu, you can specify from 3 population features: bivariate normal, uniform in the x -values, or conditional on the observed x -values. After you make your selection again press the **Create Population** button. Uncheck the Design Population box to return to the original sample data.

Theory-Based Inference

The user can choose a **Scenario** (one proportion, one mean, two proportions, two means) and either paste in **data** or use **summary statistics** to view descriptive and theory-based inferential statistics. With categorical data, you only need to enter either the count(s) or the sample proportion(s) and the other will be solved for. For quantitative data, if only summary statistics are given, the graph displays the mean $\pm$ one SD. With two groups, raw data can be pasted in as stacked or unstacked using the check box. With stacked data, use the Group Value button to specify which is the categorical variable before pressing Use Data. The group labels can be numbers or single words. (If you use group labels, they should be used in the output.) When comparing two groups, there is a toggle button below the sample graph that allows you to change the direction of subtraction.

Checking the **Test of significance** box allows specifying the direction of the alternative and after pressing **Calculate** displays the test statistic and p -value. Changing the hypothesized value in one box automatically changes it in the other. After making changes, press Calculate again. The line at the observed value is also moveable with your mouse to see how the test statistic and p -value are impacted. With a single categorical variable, you can check the box for a continuity correction (**cont corr.**) With quantitative data, the pdf displayed is for the t distribution (that's why the variable name is in quotes, see below the axis label for the t -scale). The blue vertical line is moveable, updating the test statistic and p -value output. Checking the **Confidence interval** box, specifying a confidence level, and pressing **Calculate CI** will display the confidence interval (as well or instead).

Power Simulation

You can enter two different process probabilities and generate distributions of number of successes or proportions of successes for each, for the same sample size (e.g., a null value and an alternative value). (Right now the second distribution does not appear until you check the **Show alternative** box.)

You can then choose whether you want to count observations by inputting a rejection region value (toggling the direction) or a level of significance (the direction is chosen based on the relationship between the null and alternative values). The applet reports the proportion in the specified region for each distribution. There are also options for looking at two-tailed probabilities, exact binomial probabilities, or normal approximation. You can proceed directly to the theoretical distributions and the rejection region is determined by the probability distribution rather than the simulation. The **Summary Stats** checkbox will also display the mean and SD values for both distributions. The **Show previous** checkbox adds the distribution under the previous input settings.

Dolphin Study

This is a specialized version of the two-way table applet with tables hard coded for the specific study. The animations involve blue and green cards (with a dark blue line across the top indicating original membership in “Group A”). When cards are redistributed, both groups are displayed.

Sampling Words

This proceeds to the One Mean applet but with the Gettysburg Address data (Ch. 2) as the default. The population is the 268 words in the Gettysburg Address and the variables are the length of the word, whether or not the word contains the letter e, and whether or not the word is a noun. The selected variable and its characteristics display in the population graph.

Simulating Confidence Intervals

The user can explore several confidence interval methods for one proportion or one mean. Continuing to press **Sample** will update the **Running Total**. Intervals that capture the specified parameter value are green, those that don't are red. Clicking on an interval will display its midpoint and endpoints. The distribution of the most recently selected sample and the distribution of the sample statistics are displayed. In the latter, the statistics corresponding to intervals that do not contain the parameter are highlighted in red. Pressing **Sort** will rearrange the displayed intervals by the value of the midpoint. Changing the confidence level (**Conf level**) and pressing **Recalculate** will use the same samples but change the intervals to help visually demonstrate changes in width.

For One Proportion, the Wald interval is the one-sample z-interval using the sample proportion in the standard error, the Plus Four interval adds two successes and two failures to the sample before calculating a one-sample z-interval, but only for 95% confidence. The more general Adjusted Wald procedure will allow you to change the confidence level and the Score method inverts the test statistic using the quadratic formula. Switching from Binomial to Finite Population allows the user to specify a population size, and then you can switch from Wald to the Wald with a Finite Correction to compare the coverage rates.

For One Mean, the methods are a one-sample z-interval using the inputted value of sigma, one-sample z-interval but using the values of the sample standard deviation each time, and the one-sample *t*-interval. You can now also select from different population shapes (normal, uniform, exponential) to explore robustness of the *t* intervals.

Randomizing Subjects

Demonstrates how random assignment tends to balance out variables between groups. The participants are shuffled and randomly assigned to the two treatment groups and the difference in means (heights or x-factor) or proportions (gender or balance gene) is computed. Repeating this process generates a randomization distribution that should be centered at about zero. The “balance gene” is used in a context of seeing whether different recovery strategies can help people recover from a fall. The x-factor is a second quantitative variable but with noticeably more variability than the heights. Rand after Block performs the random assignment separately for males and females (more of a generalized block design).

This [version](#) changes the subjects' names and uses BMI rather than x-variable.

Random Numbers

Allows the user to specify a range of integers to randomly sample from. You can control how many observations are sampled and how many replications. You can choose to sample with or without replacement. You can choose whether or not to sort the results.

Correlation Guessing Game

User specifies the number of observations and then generates random pairs of quantitative data (0-100) by pressing **New Sample**. You can then edit the data or paste in your own data set (use one word variable names) by checking the **Edit/Paste Data** button to open a data window.

Student should then enter a guess for the correlation coefficient and then press **Check Guess** to see the actual correlation coefficient. This can be repeated numerous times. To see a record of their guesses and the actual values, check the **Track Performance** box. This displays a

scatterplot (and correlation coefficient) between the guesses and actual values, a scatterplot between the errors in their guesses and the actual values, and a timeplot of the errors by the trial number.

Monty Hall Game

As on the popular TV show, there is a desirable prize (car) behind one of the doors and a less desirable prize (goats) behind the other two. The user clicks on a door and one of the goats is revealed. The user can then stay with the original door or switch to the other closed door. If the strategy is applied repeatedly (can uncheck the Animate box), the applet tracks the cumulative relative frequency of wins for each strategy.

Some Other Applets to Consider (under development)

- [Guess the p-value](#)
- [Two-sample bootstrapping](#)
- [Sampling from two processes](#) - categorical
- [Sampling from two populations](#) - quantitative
- [Sampling from two probability distributions](#) - quantitative

Intermediate applets [Guide \(2021\)](#)