

Introduction

The solution to the game predicts the outcomes of experimental subjects playing the game in a classroom laboratory. The conduct of the experiment is the data generation process. Measuring and interpreting the deviations observed in data from the predictions of the theory is the province of statistical inference. The tools of statistics provide a metric for accepting or rejecting the behavioral hypotheses put forward. The more closely the experimental outcomes mimic the predictions, the greater confidence we place in the theory that yields them. This is the reason for forming statistics to test and evaluate whether the behavior of the experimental subjects conforms to the solution derived for the game. When there is a unique solution to the game, it provides the benchmark data generating process against which all violations should be considered. When there are multiple solutions to the model, the process of testing them is complicated by the fact that the experimental subjects playing in a game together might not coordinate their behavior around the same solution. In any event, we might be more concerned about certain features of the data than others, and should accordingly formulate tests that conform as closely as possible to the questions we seek to answer.

This text routinely applies the theory of probability and statistics to experimental data to evaluate the solutions derived for the games. This chapter has two purposes, to provide an introduction to displaying, summarizing and analyzing data collected from experimental sessions, and to serve as a reference for the remaining chapters in the book. It reviews hypotheses tests and estimation methods for the parameters of interest implied by the solution(s) to the game. These hypotheses tests and parameters are the basis for our enquiry into how the subjects for experiment played the games, focusing especially on how closely actual play conforms to the predictions of the solution. In this section we use several simple games to demonstrate how this exercise can be conducted.

Data has different forms depending on the type of the design of the experiment. It is convenient to divide the chapter into two parts that reflect the characteristics of data. There are two types of data, categorical data or numerical data. Many types of experiments result in qualitative rather than quantitative response variables, so that responses can be classified but not quantified. Data from those experiments consist of count or number of observations falling in each of the response categories included in the experiment. numerical data, on the other hand, can be ordered.

Categorical Data

We begin with our discussion of an experiment.

1.1. A Description of the Experiment

Many experiment result in enumerative (or count) data. For example, based on the decisions that subjects made in the experiment presented in Figure 1 there are three

possibilities how the game will end: 1) In entrepreneur decides to ignore idea and continue office work the game will end in the box numbered 3; 2) if the entrepreneur decided to expend effort seeking funding for project then venture capitalist has a chance to respond and if venture capitalist selected not to fund the project then the number associated with this decision is 4 in Figure 1; 3) if venture capitalist decided to fund the project then it all depends what happens with the project (i.e. 0.5 chance the project falls and 0.5 chance the project is successful). We are only interested in subjects decisions and not how the nature draws between success and failure. Therefore we are interested how many times number 5 is visited.

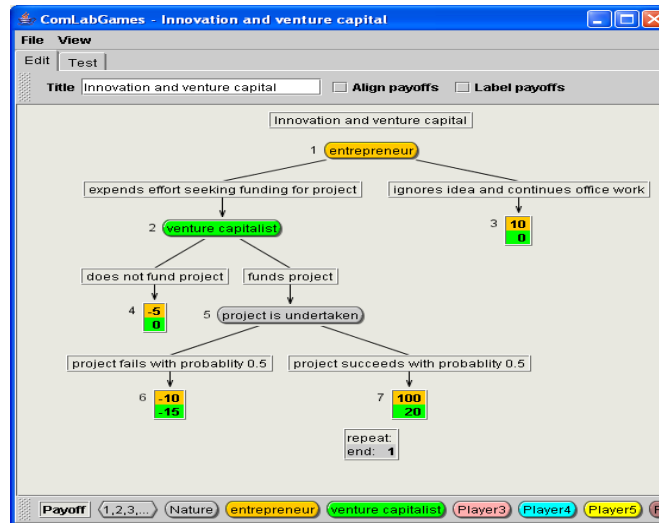


Figure 2.1

Or we might be interested in studying the reaction of responders to the proposals made by the players in the bargaining game. The responder can accept or reject the proposal. In this case the experiment will yield two counts, indicating the number of subjects falling in each of the reaction classes.

Write more examples.....

The illustration in the preceding paragraph exhibits the following characteristics that define a multinomial experiment.

The Characteristics of a Multinomial Experiment

1. The experiment consists of n identical trials (rounds).
2. The outcome of each trial (round) falls into one of the k classes or cells.
3. The probability that the outcome of a single trial will fall in a particular cell, for example cell i is p_i ($i=1, 2, \dots, k$)

Displaying Categorical Data:

The Characteristics of a Multinomial Experiment

1. The experiment consists of n identical trials (rounds).
2. The outcome of each trial (round) falls into one of the k classes or cells (i.e. results at the end of the tree).
3. The probability that the outcome of a single trial will fall in a particular cell, for example cell i is p_i ($i=1, 2, \dots, k$) and remains the same from trail to trail. Note that $0 \leq p_i \leq 1$ for all i , and

$$p_1 + p_2 + p_3 + \dots + p_k = 1$$

4. Trial are independent.
5. The experimenter is interested in $n_1, n_2, n_3, \dots, n_k$, where n_i ($i = 1, 2, \dots, k$) is equal to the number of trails in which the outcome falls in cell i . Note that $n_1 + n_2 + n_3 + \dots + n_k = n$.

A Test of Hypothesis Most commonly used in the Experiments using Categorical Data

Estimating a Binomial Proportion

When choices in the experimental design are describe (i.e. for example representation of a game in the extensive or in the strategic form) and the solution of the game is unique then this is an example of the binominal experiment. Estimating the proportion of subjects that are playing according to the equilibrium prediction is a problem that requires the estimation of a binomial parameter p . The estimator p , denoted by the simbol \hat{p} , is the total number x of successses (i.e. total number of times subjects played according to the equilibrium prediction) divided by the total number n of trials:

$$\hat{p} = \frac{x}{n}$$

where x is the number of successes in n trials. By "best" we mean that \hat{p} is unbiased and possesses a smaller variance than other possible estimators.

The estimator \hat{p} possesses a sampling distribution that can be approximated by a normal distribution because of the Central Limit Theorem. It is an unbiased estimator of the population proportion p , with mean and standard deviation:

Mean and Standard Deviation of \hat{p}

$$E(\hat{p}) = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

Point Estimator for p

Estimator: $\hat{p} = \frac{x}{n}$

Margin of error: $1.96\sigma_{\hat{p}} = 1.96\sqrt{\frac{pq}{n}}$

Estimated margin of error: $1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$

The corresponding large-sample confidence interval with confidence coefficient $(1-\alpha)$ is:

A $(1-\alpha)$ 100% Confidence Interval for p

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Example:

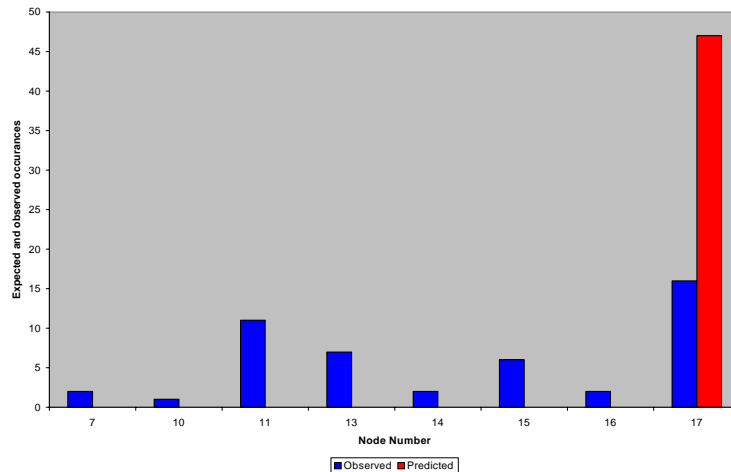


Figure 1: χ^2 statistics for Armageddon in extensive form

A random sample of $n=48$ games produced $x=16$ plays that ended in terminal node 17 (solution to this game). Estimate the fraction of the experimental population ending in the node that is predicted by the theory and find the estimated margin error.

The point estimate is:

$$\hat{p} = \frac{17}{48} = 0.362$$

and the margin of error is

$$1.96\sigma_{\hat{p}} = 1.96\sqrt{\frac{pq}{n}} \approx 1.96\sqrt{\frac{0.362 * 0.638}{48}} = 0.137$$

A 95% confidence interval for p would be

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} \text{ or } 0.362 \pm 1.96 * 0.0049$$

Thus, we would estimate that p lies in the interval 0.225 to 0.499 with confidence coefficient 0.95. Notice that 0.362 ± 0.137 includes proportions that are smaller than 0.5. Therefore we can conclude that subjects did not play according to the theoretical prediction.

Test if the estimated \hat{p} is significantly different from the predicted value of 1. Check this!!!!!!!

Chi- Square Test

The χ^2 test gives a goodness-of-fit measure for the observed and theoretical probability distribution (discrete variables) or densities (continuous). General formulae is:

$$\chi^2 = \sum_{i=1}^k \frac{[n_i - E(n_i)]^2}{E(n_i)}$$

for discrete variables. n_i means number of observed occurrences in a class and $E(n_i)$ means the number of predicted occurrences.

For continuous variables the χ^2 test has the following representation:

$$\chi^2 = \sum_{i=1}^k \frac{[n_i - np_i]^2}{np_i}$$

We also have to determine the degrees of freedom associated with the appropriate chi-square distribution. Furthermore, we must determine whether the rejection region for the test is one-tailed or two-tailed (check when each).

To determine the appropriate number of degrees of freedom we state the principle involved in order to understand why the number of degrees of freedom changes with various applications. The principle states that the appropriate number of degrees of freedom (d.f.) equals the number of cells k less 1 d.f. for each independent linear

restriction placed on the observed cell counts. For example, one linear restriction is always present because the sum of the cell counts must equal n ; that is:

$$n_1 + n_2 + n_3 + \dots + n_k = n$$

Other restrictions will be introduced for some application because of the necessity for estimating unknown parameters required in the calculation of the expected cell frequencies or because of the method in which the experiment was conducted and the data collected.

The χ^2 test gives a goodness-of-fit measure for the observed and theoretical probability distribution (discrete variables) or densities (continuous). General formulae is:

$$\chi^2 = \sum_{\text{classes}} \frac{n_{\text{obs}} - n_{\text{exp}}}{n_{\text{exp}}}$$

for discrete variables. n_{obs} means number of observed occurrences in a class and n_{exp} means the number of predicted occurrences.

For continuous variables the χ^2 test has the following representation:

$$\chi^2 = \sum_{\text{classes}} \frac{n_{\text{obs}} - np(\text{data in class})}{np(\text{data in class})}$$

Many times we count the number of occurrences of an event. For example consider the following two games. The first game is represented by its extensive form and depicted in Figure 3.1.

To test whether the groups are behaving the same way. Considering the following contingency table constructed from the outcomes, we seek to test whether both groups of subjects behaved the same way.

	Terminal node 3; Up/right or down/right	Terminal node 4; Up/left	Terminal node 6 and Terminal node 7; Down/left	Total
Extensive	14	6	2 + 3 = 5	25
Strategic	1	2	22	25
Total	15 (0.3)	8 (0.16)	27 (0.54)	50

Let H_0 be the hypothesis that the proportions in the three possible ways the games can end is the same in extensive and in the strategic form. Of 50 completed sessions 15/50 or 30% ended up in Terminal node 3 (i.e. up/right or down/right); 8/50, or 16% ended in Terminal node 4 and 27/50, or 54% ended either in terminal node 6 or 7. There are 25 completed sessions in extensive form. Therefore 30% of 25 sessions or

7.5 sessions should end in terminal node 3. Similarly 16% of 25 or 4 should end in Terminal node 4 and 54% or 13.5 sessions should end in node 6 or 7. Under H_0 the same proportion should hold for the strategic form.

To determine whether the differences are more than could reasonably be due to chance alone, we use a chi-square test. The test statistic is:

$$\begin{aligned}\chi^2 &= \sum [\text{Observed-Expected}]^2/\text{Expected}] \\ &= (14 - 7.5)^2/7.5 + (6 - 4)^2/4 + (5 - 13.5)^2/13.5 + (1 - 7.5)^2/7.5 \\ &+ (2 - 4)^2/4 + (22 - 13.5)^2/13.5 \\ &= 23.97\end{aligned}$$

The number of degrees of freedom is computed as follows: Let r = the number of rows in the table. There are two: extensive form and strategic form. Let c = the number of columns. There are three: Terminal node 3, terminal node 4 and terminal node 6 or 7. The number of degrees of freedom is $(r-1)(c-1) = (2-1)(3-1) = 2$. For $\alpha = 0.005$, CDF gives 7.88. Our observed value is much larger than this. We conclude that subjects playing extensive form played differently from the subjects who played the strategic form.

The sample for the extensive form game was further split over the first and second part of the game session. In the 12 completed sessions in the first half:

- 1 terminated at node 3
- 6 at node 4
- 2 at node 6
- 3 at node 7

The test follows the same procedure.

A Test of Hypothesis Concerning Specified Cell Probabilities

The simplest hypothesis concerning the cell probabilities is one that specifies numerical values for each cell. For example, in the Armageddon experiment in the extensive form, subjects deviated from the predicted behavior and have not selected the choices that would have resulted in node number 17. Now we wish to test the hypothesis that subjects will equally likely choose any of the other terminal nodes (i.e. approximately 1/8 of the time). Therefore

$$H_0 : p_1 = p_2 = \dots = p_8 = \frac{1}{8}$$

versus

$$H_a : \text{at least one } p_i \text{ is different from } \frac{1}{8}$$

where p_i is the probability that a subjects will end in nodes 7, 10, 11, 13, 14, 15 and 16.

There were $n=48$ of cases that subjects ended in one of the eight nodes and the observed cell frequencies were the following:

finished later

Contingency Tables

A problem frequently encountered in the analysis of count data concerns the independence of two methods of classifying observed events. For example, we wish to classify results of the experiment on Innovation and Venture Capital - first according to the terminal node that it was reached and, second, according to the presentational form the subjects were given in the experimental session (i.e. in our case we are comparing the results from the extensive form game and strategic form game). If the proportions of ending the game in different nodes are constant from one presentation to the other, then classification by terminal nodes (choice selection) is independent of the classification by the presentation given to the subjects. On the other hand, if the proportions of ending in different terminal nodes (i.e. different selection of choices) vary from one presentation to the other, then classification by terminal nodes is contingent upon the presentation classification, and the classifications are dependent. In investigating whether one method of classification is contingent upon another, we display the data by using a cross-classification in an array called a **contingency table**.

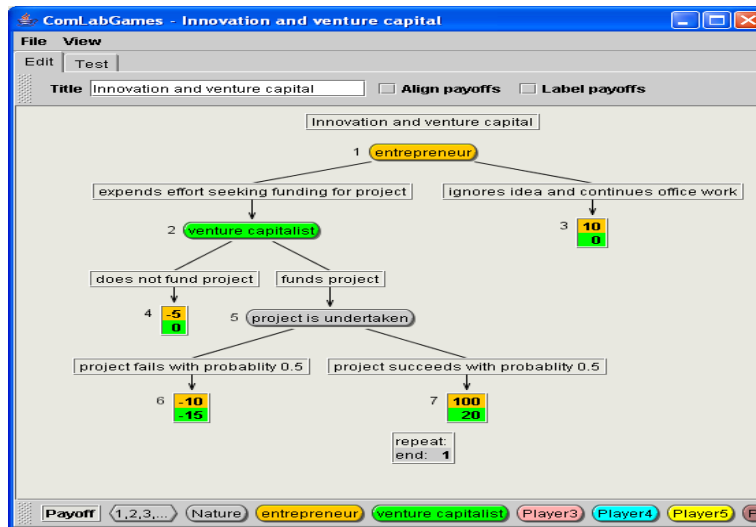


Figure 2.1: Extensive form game representation

The second game is represented by its strategic form and depicted in Figure 2.2.

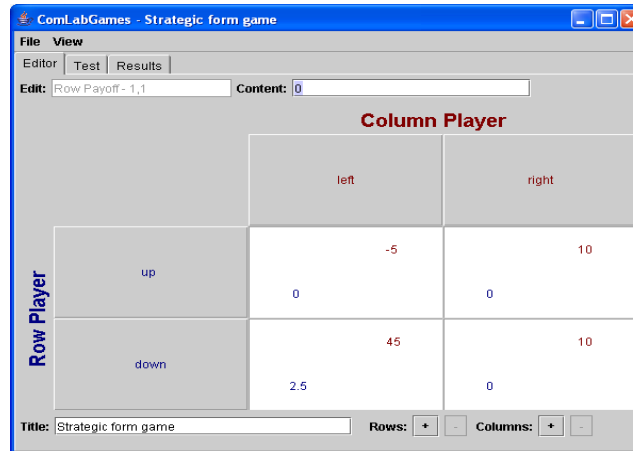


Figure 2.2: The strategic form representation of the game

The game depicted in Figure 2.1 is the same game as the one depicted in Figure 2.2. In the strategic form representation the entrepreneur is the column player, venture capitalist is the row player and they both maximize their expected income. Then there are 2 strategies for the entrepreneur (expend effort = left, and ignore idea = right) and 2 strategies for the venture capitalist (do not fund = up and fund = down). The payoffs are then filled in appropriately (where the payoff from down/left for example is computed from $\frac{1}{2}(-10, -15) + \frac{1}{2}(100, 20)$).

Two groups of subjects drawn from the same population pool were selected to participate in an experiment. The first group of subjects played the extensive form game depicted in Figure 3.1. In the 25 completed sessions:

In the 25 completed sessions:

- 14 terminated at node 3,
- 6 at node 4
- 2 at node 6
- 3 at node 7

The second group played the strategic form game depicted in Figure 3.2. In the 25 completed sessions:

- 2 at up/left
- 0 at up/right
- 22 at down/left
- 1 at down/right

A total of $n=50$ decisions were recorded and the three possible ways the game can end: Terminal node 3 or up/right or down/right, terminal node 4 or up/left, terminal node 6 and terminal node 7 or down/left. At the same time, each decision was identified according to the game form subjects played. These counts are presented as a contingency table in Table 2.1 (NOTE: Numbers in parantheses are the expected cell frequencies.)

	Terminal node 3 Up/Right or down/right	Terminal node 4 up/left	Terminal node 6 and 7 down/left	Total
Extensive form	14 (7.5)	6 (4)	2+3=5 (13.5)	25
Strategic form	1 (7.5)	2 (4)	22 (13.5)	25
Total	15	8	27	50

Table 2.1: Contingency table

Let p_3 be the unconditional probability that the decision of subjects will ebd in terminal node 3. Similarly define p_4 and $p_{6,7}$ as probabilities observing the two other types of outcomes. Then these probabilities, which we will call the column probabilities of Table 2.1 will satisfy the requirement

$$p_3 + p_4 + p_{6,7} = 1$$

In the like manner, let p_i ($i = E$ and S) row probability that the outcome come from the extensive form and the strategic form respectively where

$$p_E + p_S = 1$$

Then if the two classifications are independent of each other, a cell probability will equal the product of its respective row and column probabilities in accordance with the multiplicative law of probabilities.

For example, the probability that a terminal node 3 will be selected when subjects were given to play the game in the extensive form is $(p_3)(p_E)$. Thus we observe that the numerical values of the cell probabilities are unspecified in the problem under consideration. The null hypothesis specifies only that each cell probability will equal the product of its respective row and column probabilities and therefore imply independence of the two representatins of the game (i.e. classifications). The alternative hypothesis is that this equality does not hold for at least one cell. In order to test the hypothesis we must estimate the row and column probabilities so that we can estimate the expected cell frequencies. The procedures for obtaining the estimates is called the method of maximum likelihood and the method of minimum chi-square.

The maximum likelihood estimator of a column probability will equal the column total divided by $n=50$. If we denote the total of column j as c_j , then

$$p_3 = \frac{c_3}{n} = \frac{15}{50} \quad p_4 = \frac{c_4}{n} = \frac{8}{50} \quad p_{6,7} = \frac{c_{6,7}}{n} = \frac{27}{50}$$

Similarly, the row probabilities p_E and p_S can be estimated using the row totals r_E and r_S :

$$p_E = \frac{r_E}{n} = \frac{25}{50} \quad p_S = \frac{r_S}{n} = \frac{25}{50}$$

Denote the observed frequency of the cell in row i and column j of the contingency table n_{ij} . Then the estimated expected value of n_{11} will be

$$\hat{E}(n_{11}) = n[\hat{p}_E \hat{p}_3] = n\left(\frac{r_E}{n}\right) \left(\frac{c_3}{n}\right) = \frac{r_E c_3}{n}$$

where $(\hat{p}_E \hat{p}_3)$ is the estimated cell probability. Similarly, we can find the estimated expected value for any other cell. Thus, we see that the estimated expected value of the observed cell frequency n_{ij} for a contingency table is equal to the product of its respective row and column totals divided by the total frequency; that is,

Estimated Expected Cell Frequency

$$\hat{E}(n_{ij}) = \frac{r_i c_j}{n}$$

where

$$r_i = \text{total for row } i$$

$$c_j = \text{total for column } j$$

The estimated expected cell frequencies for our example are shown in parentheses in Table 2.1. We can now use the expected and observed cell frequencies shown in Table 2.1 to calculate the value of the test statistic

$$\begin{aligned} \chi^2 &= \sum_{j=1}^3 \sum_{i=1}^2 \frac{[n_{ij} - \hat{E}(n_{ij})]^2}{\hat{E}(n_{ij})} \\ &= \frac{(14 - 7.5)^2}{7.5} + \frac{(6 - 4)^2}{4} + \dots + \frac{(22 - 13.5)^2}{13.5} \\ &= 23.97 \end{aligned}$$

The next step is to determine the appropriate number of degrees of freedom associated with the test statistic. The degrees of freedom associated with a contingency table having r rows and c columns will always equal $(r-1)(c-1)$. Thus, for our example, we will compare χ^2 with the critical value of χ^2 with $(r-1)(c-1) = (2-1)(3-1) = 2$ d.f.

The number of degrees of freedom associated with the χ^2 statistic equals the number of cells ($k=rc$) less one degree of freedom for each independent linear restriction placed on the observed cell frequencies. The total number of cells for the data of Table 2.1 is $k=6$. From this we subtract 1 d.f. because the sum of the observed cell frequencies must equal n ; that is,

$$n_{11} + n_{12} + \dots + n_{23} = 50$$

In addition, we used the cell frequencies to estimate two of the three column probabilities. Note that the estimate of the third column probability will be determined once we have estimated p_3 and p_4 because

$$p_3 + p_4 + p_{6,7} = 1$$

We lose $(c-1)=2$ d.f. for estimating the column probabilities. We used cell frequencies to estimate $(r-1)=1$ row probabilities and, therefore we lose $(r-1)=1$ additional degrees of freedom. The total number of degrees of freedom remaining will

be

$$d.f. = 6 - 1 - 2 - 1 = 2$$

In general, we see that the total number of degrees of freedom associated with an rc contingency table is

$$\begin{aligned} d.f. &= rc - 1 - (c - 1) - (r - 1) \\ &= rc - c - r + 1 = (r - 1)(c - 1) \end{aligned}$$

If we use $\alpha = 0.05$. We will reject the null hypothesis that the two representations are independent (i.e. subjects in the extensive form behave the same way as subjects given the strategic form) if $\chi^2 > 7.88$. Since the value of test statistic, $\chi^2 = 23.97$, exceeds the critical value of χ^2 , we reject the null hypothesis. The data present sufficient evidence to indicate that the proportion of various ways the game can end varies between the two representations.

Most statistical packages include a program for analyzing data contained in an rc contingency table.

Other applications

Time-Dependent Multinomials

Multidimensional Contingency Tables

Assumptions

The following assumptions must be satisfied if χ^2 is to have approximately a chi-square distribution and consequently, the test described in this chapter are to be valid.

Assumptions

1. The cell counts n_1, n_2, \dots, n_k satisfy the conditions of a multinomial experiment (or a set of multinomial experiments created by restrictions on row and column totals).
2. The expected values of a cell counts should equal or exceed 5.

Assumption 1 must be satisfied. The chi-square goodness-of-fit tests, of which these tests are special cases, compare frequencies with expected frequencies and apply only to data generated by a multinomial experiment.

The larger the sample size n , the closer the chi-square distribution will approximate the distribution of χ^2 .

Nonparametric Statistics

The Mann-Whitney U Test: Independent Random Samples

The t test for comparing the population means is a test to detect differences in the location of two normal population frequency distributions. The Mann-Whitney U test is a nonparametric alternative to this test. It is used when we have doubts that the assumptions of normality and/or equal variances required for the Student's t test are satisfied. The Mann-Whitney U test (and all of the tests that follow in this section) is

based on an analysis of the ranks of the sample observations.

Assume that you have independent random samples of sizes n_1 and n_2 from two populations, 1 and 2. The first step in finding the Mann-Whitney U statistic is to rank all $(n_1 + n_2)$ observations in order of magnitude, assigning 1 to the smallest observation, a 2 to the second smallest, and so on. Ties in the observations can be handled by averaging the ranks that would have been assigned to the tied observations and assigning this average to each. Then calculate the sums of the ranks, T_1 and T_2 for the two samples.

Write the example

The formula for the Mann-Whitney U statistic can be given in terms of T_1 and T_2 ; one value of U will be larger than the other, but the sum of two U values will always be equal $n_1 n_2$. The formulas for the two values of U : U_1 and U_2 are given below:

Formulas for the Mann-Whitney U Statistic

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - T_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_2$$

where

n_1 = number of observations in sample 1

n_2 = number of observations in sample 2

T_1 and T_2 are the rank sums of samples 1 and 2, respectively. (Note: It can be shown that $U_1 + U_2 = n_1 n_2$ and $T_1 + T_2 = \frac{n(n+1)}{2}$, where $n = n_1 + n_2$.)

U_1 will be small when T_1 is large. This will likely occur when population 1 distribution of measurements is shifted to the right of the population 2 distribution of measurements. To conduct a one-tailed test to detect a shift in distribution 1 to the right of distribution 2, you will reject the null hypothesis of "no difference in the population distributions" if U_1 is less than some specified value U_0 ; that is, you will reject H_0 for small values of U_1 . Similarly, to conduct a one-tailed test to detect a shift of distribution 2 to the right of distribution 1, you would reject H_0 if U_2 is less than some specified value, say U_0 .

Table **X** in Appendix **X** gives the probability that an observed value of U will be less than some specified value U_0 . This is the value of α for a one-tailed test. To conduct a two-tailed test- that is, to detect a shift in the population distributions for the measurements in either direction - we use U , the smaller of U_1 or U_2 as the test statistic and reject H_0 for $U < U_0$. The value of α for the two-tailed test will be double the tabulated value given in table **X** in Appendix **X**.

To see how to locate the rejection region for the Mann-Whitney U test, suppose that $n_1 = 4$ and $n_2 = 5$ (see table X). Across the top of Table X, you see the value of n_1 . Values of U_0 are shown down the left side of the table. The entries give the probability that U will assume a small value (i.e. probability that $U \leq U_0$). Since, for this example, $n_1 = 4$, we move across the top of the table to $n_1 = 4$. Move to the second row of the table corresponding to $U_0 = 2$ for $n_1 = 4$. The probability that U will be less or equal to 2 is found as 0.0317. Similarly moving across the row for $U_0 = 3$, you see that the probability that U is less or equal to 3 is 0.0556. If you conduct a one-tailed Mann-Whitney U test with $n_1 = 4$ and $n_2 = 5$ and would like α to be near 0.05, you would reject the null hypothesis of equality of population relative frequency distributions when $U \leq 3$. The probability of a type I error for the test would be $\alpha = 0.0556$. If you use this same rejection region for a two-tailed test, that is, $U \leq 3$, α would be double the tabulated value, or $\alpha = 2(0.0556) = 0.1112$.

When applying the test to a set of data, you may find that some of the observations are of equal value. Recall that ties in the observation can be handled by averaging the ranks that would have been assigned to the tied observations and assigning this average to each. For example, if three observations were tied and were due to receive ranks 3, 4, and 5, we would assign the rank of 4 to all three. The next observation in the sequence would receive the rank of 6, and ranks 3 and 5 would not appear. Similarly if two observations were tied for ranks 3 and 4, each would receive a rank of 3.5, and ranks 3 and 4 would not appear.

The Mann-Whitney U Test

1. *Null hypothesis: H_0* : The population relative frequency distribution 1 and 2 are identical.
2. *Alternative Hypothesis: H_a* : The two population relative frequency distributions are shifted with respect to their relative locations (a two-tailed test). Or H_a : The population relative frequency distribution 1 is shifted to the right of the relative frequency distribution for population 2 (a one-tailed test).
3. *Test Statistic*: For a two-tailed test, use U , the smaller of

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - T_1$$

and

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_2$$

where T_1 and T_2 are the rank sums for samples 1 and 2.

Example 2.3

In a bargaining game subjects have to decide how much they will offer to the opponent in the range from 0 to 10. One experiment conducted involved only male subjects pool and the other experiment involved the female subject pool. 10 observations were reported for each experiment:

Male subject pool	Female Subject pool
1.21 (2)	1.49 (15)
1.43 (12)	1.37 (7.5)
1.35 (6)	1.67 (20)
1.51 (17)	1.50 (16)
1.39 (9)	1.31 (5)
1.17 (1)	1.29 (3.5)
1.48 (14)	1.52 (18)
1.42 (11)	1.37 (7.5)
1.29 (3.5)	1.44 (13)
1.40 (3.5)	1.53 (19)
Rank sum $T_1 = 85.8$ $T_2 = 124.5$	

Suppose we choose a value for α near 0.05. Then we can find U_0 by consulting the portion of Table X, Appendix..., corresponding to $n_2 = 10$. The probability, $P(U \leq U_0)$, nearest 0.05 is 0.0526 and corresponds to $U_0 = 28$. We will reject if $U_2 \leq 28$.

Calculating U_2 , we have

$$\begin{aligned} U_2 &= n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_2 \\ &= (10)(10) + \frac{(10)(11)}{2} - 124.4 \\ &= 30.5 \end{aligned}$$

As you can see, U_2 is not less than $U_0 = 28$. Therefore, we cannot reject the null hypothesis. At the $\alpha = 0.05$ level of significance, there is not sufficient evidence to indicate that offers made by females are larger than the offers made by male subject pool.

A simplified large-sample test ($n_1 \geq 10$ and $n_2 \geq 10$) can be obtained by using z statistic (see subsection z). When the population distributions are identical, it can be shown that the U statistic has expected value and variance:

$$E(U) = \frac{n_1 n_2}{2} \text{ and } \sigma_U^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

The distribution of

$$z = \frac{U - E(U)}{\sigma_U}$$

tends to normality with mean 0 and variance equal to 1 as n_1 and n_2 become large. This approximation would be adequate when n_1 and n_2 are both greater than or equal to 10. Thus, for a two-tailed test with $\alpha = 0.05$, we would reject the null hypothesis if

$|z| \geq 1.96$.

Observe that by using the z statistic, you will reach the same conclusions as when using U test for the example 2.3. Thus,

$$z = \frac{30.5 - [(10)(10)/2]}{\sqrt{[(10)(10)(10 + 10 + 1)]/12}} = \frac{30.5 - 50}{\sqrt{2100/12}} = -1.47$$

For a one-tailed test with $\alpha = 0.05$ located in the lower tail of the z distribution, we reject the null hypothesis if $z < -1.645$. Since $z = -1.47$ does not fall in the rejection region, we reach the same conclusion as the exact U test for the example 2.3.

The Mann-Whitney U Test for Large Samples: $n_1 \geq 10$ and $n_2 \geq 10$

write it down

The Wilxon Signed-Rank Test for a Paired Experiment

Calculation of the Test Statistic for the Wilcoxon Signed-rank Test

1. Calculate the difference ($x_A - x_B$) for each of the n pairs. Differences equal to 0 are eliminated and the number of pairs, n , is reduced accordingly.
2. Rank the absolute values of the differences, assigning 1 to the smallest, 2 to the second smallest, and so on. Tied observations are assigned the average of the ranks that would have been assigned with no ties.
3. Calculate the rank sum for negative differences and label this value T^- . Similarly calculate T^+ , the rank sum for the positive differences.

For a two-tailed test, we use the smaller of these two quantities as a test statistic to test the null hypothesis that the two population relative frequency histograms are identical. The smaller the value of T , the greater will be the weight of evidence favoring rejection of the null hypothesis. Therefore we will reject the null hypothesis if T is less than or equal to some value, T_0 .

For one-sided test to test that distribution 1 is shifted to the right of distribution 2, we should use the rank sum T^- of the negative differences and reject the null hypothesis for small values of T^- , $T^- \leq T_0$. If we wish to detect a shift of distribution 2 to the right of distribution 1, we use the rank sum T^+ of the positive differences as a test statistic and reject for small value of T^+ , $T^+ \leq T_0$.

The probabilities that T is less than equal to some value T_0 has been calculated for a combination of sample sizes and values of T_0 . These probabilities, given in Appendix X can be used to find the rejection region for the T test.

Example 2.4 Test an hypothesis of no difference in population distributions of deviation of bids for the paired difference in the auction experiment that included the comparison of bids in Second price sealed bid auction and English auction. Six different subjects participated in each auction experiment. The valuations were

randomly drawn from uniform distribution with support [50, 200] but the valuations that were drawn in one experiment were randomly assigned in the second experiment.

	Valuation 1	Valuation 2	Valuation 3	Valuation 4
Bid in second price sealed bid auction (x_1)	135	102	98	141
Bid in English auction (x_2)	129	120	112	152
Difference ($x_1 - x_2$)	6	-18	-14	-11
Rank	2	5	4	3

Table

The null hypothesis to be tested is that the two population frequency distributions of bids are identical. The alternative hypothesis, which implies a two-tailed test, is that the distributions are different.

Because the amount of data is small, we will conduct the test using $\alpha = 0.10$. From table X, the critical value of T for a two-tailed test, $\alpha = 0.10, T_0 = 2$. Hence, we reject H_0 if $T \leq 2$.

The difference $x_1 - x_2$ are calculated and ranked according to their absolute value in the table. The sum of positive ranks is $T^+ = 2$, and rank sum of the negative ranks is $T^- = 19$. The test statistic is the smaller of these two rank sums, or $T = 2$. Since $T = 2$ falls in the rejection region, we reject H_0 and conclude that the two population frequency distributions of deviations from the predicted bids differ.

Although Table X in Appendix is applicable for values of n (number of data pairs) as large as $n = 50$, it is worth noting that T^+ , like the Mann-Whitney U , will be approximately normally distributed when the null hypothesis is true and n is large (25 or more). This allows us to construct a large-sample z test, where

$$E(T) = \frac{n(n+1)}{4}$$

$$\sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$$

Then the z statistic

$$z = \frac{T^+ - E(T^+)}{\sigma_{T^+}} = \frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

can be used as a test statistic. Thus, for a two-tailed test and $\alpha = 0.05$, we could reject the hypothesis of "identical population distributions" when $|z| \geq 1.96$.

A Large-Sample Wilcoxon Signed-Rank Test for a Paired Experiment: $n \geq 25$

1. *Null Hypothesis: H_0* : The population relative frequency distributions 1 and 2 are identical.

2. *Alternative Hypothesis: H_a* : The two population relative frequency distributions differ in location (a two-tailed test). Or H_a : the population 1 relative frequency distribution is shifted to the right (left) of the relative frequency distribution for

population 2 (a one-tailed tets).

$$3. \text{ Test Statistic: } z = \frac{T^+ - [n(n+1)/4]}{\sqrt{[n(n+1)(2n+1)]/24}}$$

4. *Rejection region:* Reject H_0 if $z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$ for a two-tailed test. For a one-tailed test, place all of α in one tail of the z distribution. To detect a shift in distribution 1 to the right of distribution 2, reject H_0 when $z > z_{\alpha}$. To detect a shift in the opposite direction, reject H_0 if $z < -z_{\alpha}$.

The Kruskal-Wallis H Test for Completely Randomized Designs

This test is used to detect differences in location among more than two population distributions based on independent random sampling.

The procedure for conducting the Kruskal-Wallis H test is similar to that used for Mann-Whitney U test. Suppose we are comparing k populations based on independent random samples n_1 from population 1, n_2 from population 2,...and n_k from population k where

$$n_1 + n_2 + \dots + n_k = n$$

The first step is to rank all n observations from the smallest (rank 1) to the largest (rank n). Tied observations are assigned a rank equal to the average of the ranks they would have received if they had been nearly equal but not tied. We then calculate the rank sums T_1, T_2

Wilcoxon-Mann-Whitney type tests for first order stochastic dominance

For data which is not normally distributed or follow some other a priori known theoretical distribution function, one must use non-parametric tests. One such test is the Wilcoxon-Mann-Whitney test.

Data from two samples are combined and the data are tagged so that the batch to which they belong is recorded. The combined data set is then sorted according to the value of the data. Define R_1 as the sum of ranks held by the number of batch 1 and R

Numerical Data

sample mean different from theoretical prediction (F test for normal and chi squared for asymptotic)

vector of sample means different from theoretical predictions

difference between two means

estimating coefficients in a regression analysis

testing the null effects of playing according to the theory