Chapter 7
Hypothesis Testing

Sampling Distribution

We ask many questions during our daily activity. When we conduct research we call these daily questions, research questions. Research questions then hypothesize that certain things have occurred or will occur from experimentation. For example, we might ask, Do more men or women frequent the public library? From this question comes a hypothesis: The percentage of women is higher than the percentage of men who frequent the public library. What we need is some way of determining if the percent difference between women and men, or the probability of occurrence, is beyond what would be expected by chance.

We investigate or research variables of interest by obtaining the data and forming a sampling distribution. There are many different sampling distributions, each providing an estimate of its corresponding population parameter. We therefore infer that our sample data will provide an estimate of the population. The sampling distributions provided the basis for creating different types of statistical tests, where hypotheses about the probability of occurrence could be tested. A sampling distribution is a frequency distribution of a statistic created by taking repeated samples of a given size from a population. Consequently, we can create sampling distributions of the mean, variance, standard deviation, range, or median, as well as many other sample statistics, with each providing a sample estimate of a corresponding population parameter.

The statement, "*A statistic is to a sample as a parameter is to a population*," is a very important concept in statistics. This basic statement reflects the idea behind taking a random sample from a population, computing a statistic, and using that sample statistic as an estimate of the population parameter. Obviously, if the population parameter were known, e.g., mean or standard deviation, then we would not need to take a sample of data and estimate the population parameter.

All sample statistics have sampling distributions with the variance of the sampling distribution indicating the error in the sample statistic, i.e., the error in estimating the population parameter. When the error is small, the statistic will vary less from sample to sample, thus providing us an assurance of a better estimate of the

R. Schumacker and S. Tomek, *Understanding Statistics Using R*,
DOI 10.1007/978-1-4614-6227-9_7, © Springer Science+Business Media New York 2013

population parameter. In previous chapters, examples were provided to demonstrate that larger sample sizes yielded smaller error or variance in the sampling distribution, i.e., yielded a more reliable or efficient statistical estimate of the population parameter. For example, the mean and median are both sample statistics that estimate the central tendency in population data, but the mean is the more consistent estimator of the population mean because the sampling distribution of the mean has a smaller variance than the sampling distribution of the median. The variance of the sampling distribution of the mean is called the *standard error of the mean*. It is designated as:

$$\frac{\sigma}{\sqrt{n}}$$

and is estimated in a sample as:

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

The sampling distribution of a statistic is a function of sample size. In the formula, it is easy to see that as sample size n becomes larger, the denominator in the formula becomes larger and the standard error of the statistic becomes smaller; hence the frequency distribution of the statistic or sampling distribution has less variance. This indicates that a more precise sample estimate of the population parameter or value is achieved. This concept of standard error of a sampling distribution applies to any sample statistic that is used to estimate a corresponding population parameter.

An important concern in using sample statistics as estimators of population parameters is whether the estimates possess certain properties. Sir Ronald Fisher in the early twentieth century was the first to describe the properties of estimators. The four desirable properties of estimators are (1) *unbiased*; (2) *efficient*; (3) *consistent*; and (4) *sufficient*. If the mean of the sampling distribution of the statistic equals the corresponding population parameter, the statistic is *unbiased*; otherwise it is a biased estimator. If the sampling distributions of two statistics have the same mean, then the statistic with the smaller variance in its sampling distribution is more *efficient* (more precise or less variable) while the other statistic is a less efficient estimator. A statistic is a *consistent* estimator of the population parameter if the statistic gets closer to the actual population parameter as sample size increases. A *sufficient* statistic is one that can't be improved upon using other aspects of the sample data. If several sample statistics compete as an estimate of the population parameter, e.g., mean, median, and mode, the sample statistic that is unbiased, efficient, and consistent is a *sufficient* sample statistic estimate, while the other sample statistics are less sufficient. We are therefore interested in sample estimates of population parameters that are unbiased, efficient, consistent, and sufficient.

The sample mean (statistic) is an unbiased, efficient, sufficient, and consistent estimator of the population mean (parameter). Sample statistics however don't always possess these four properties. For example, the sample standard deviation is a biased, but consistent estimator of the population standard deviation. The sample standard deviation therefore more closely approximates the population standard deviation as sample size increases, i.e., it is a consistent estimate.

The sampling distributions of sample standard deviations are generated given varying sample sizes. The sample standard deviation is computed as:

$$S = \sqrt{\Sigma (X - Mean)^2 / (N - 1)}$$

The frequency distribution of sample standard deviations computed from repeatedly drawing samples from a population generates the sampling distributions of the sample standard deviations. A comparison of the mean of the sampling distribution of sample standard deviations to the population standard deviation will help us to determine if the sample standard deviation is a consistent estimator of the population standard deviation. This basic approach can be used to determine whether any sample statistic is a consistent estimator of a corresponding population parameter. Characteristics common to the sampling distributions based on different sample sizes are the most descriptive information we can obtain about theoretical population distributions. A statistic that estimates a parameter is unbiased when the average value of the sampling distribution is equal to the parameter being estimated. The sample standard deviation is an unbiased estimator of the population standard deviation if the mean of the standard deviation sampling distribution is equal to the population standard deviation. In this chapter, you will learn to make a judgment about whether or not the sample standard deviation is a biased or an unbiased estimator of the population standard deviation.

Many different sampling distributions can be generated with the mean of the sampling distribution being an estimator of the population parameter. For each sampling distribution of a statistic, the variance of the sampling distribution indicates how precise the statistic is as an estimator. The variance of a sampling distribution of a statistic becomes smaller as sample size increases, i.e., the standard error of the statistic. The sampling distribution of sample standard deviations is a frequency distribution of sample standard deviations computed from samples of a given size taken from a population. The variance of the sampling distribution of sample standard deviations decreases with increased sample size; thus the sample standard deviation is a consistent estimator. The mean of the sampling distribution of sample standard deviations is less than the population standard deviation; thus the sample standard deviation is a biased estimator of the population standard deviation. The error in the sample standard deviation as an estimate of the population standard deviation decreases with increased sample size.

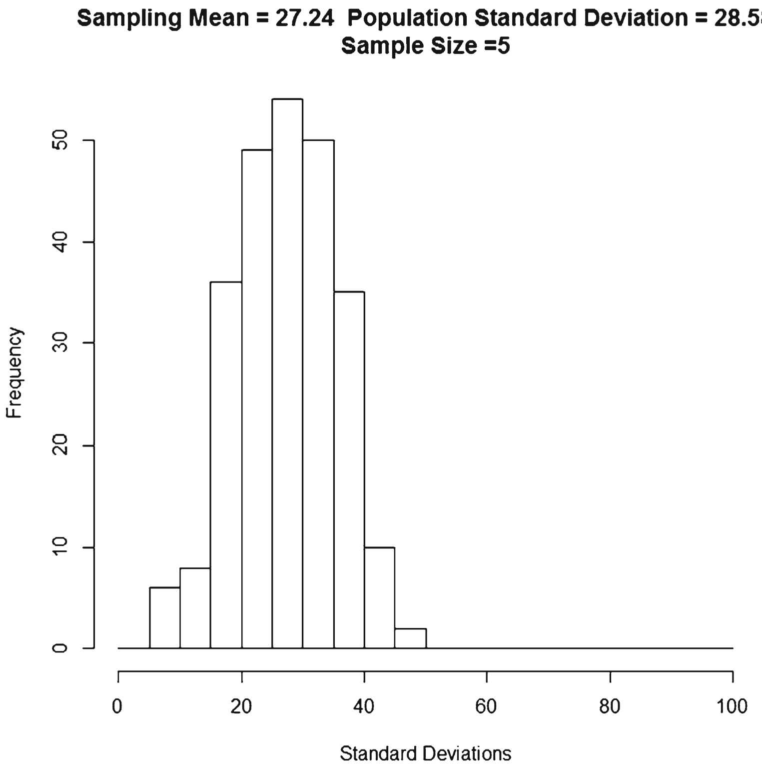
DEVIATION R Program

The DEVIATION R program produces the sampling distribution of the sample standard deviation. The program initially specifies a sample size of n=5 and then 250 replications are chosen at random from a uniform distribution between 0 and 100.

The sample standard deviation is calculated for each sample and a histogram is printed for these 250 standard deviations. The program generates a single sampling distribution for a given sample size. A different sampling distribution exists for every sample of size N. Consequently, you will need to run the program several times, each time with a different sample size, in order to determine whether the mean of the sampling distribution more closely approximates the population parameter.

The main processing loop iterates from one to the number of desired replications. For each replication, a sample of size, *SampleSize*, is drawn from a uniform population ranging from 0 to 100. The standard deviation of each sample is calculated and added to the distribution of standard deviations in the vector. After the completion of the loop, the mean of the distribution of sampling standard deviations is calculated. The mean, the sample size, and the true population standard deviation are placed into the *HistTitle* variable. Finally, a histogram of the distribution of sampling standard deviations ranging from 0 to 100 is graphed with these values printed.

DEVIATION Program Output



Deviation Exercises

1. List the four desirable properties of a sample estimate (statistic) for a population parameter.

   a. ________________
   b. ________________
   c. ________________
   d. ________________

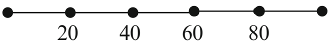
2. Compare the standard error of the mean from the following two sampling distributions.

   Note: The standard error of the mean is calculated as: $S_{\bar{X}} = \dfrac{S}{\sqrt{n}}$

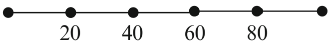
   Which sample has the smaller error? ____________________________________

   Sample 1: S = 20, N = 100 _____ Sample 2: S = 10, N = 100 ______

3. Run DEVIATION for the two sample sizes below and draw the histograms.



N = 2



N = 100

   Sampling S.D. Mean ____________ Sampling S.D. Mean _____________

   Population S.D. ________________ Population S.D. _________________

   a. Are the two graphs the same or different? Same ________ Different _______

      Why? ____

   b. Which sample size has a Sampling S.D. Mean closer to the Population S.D.?

      ____

4. Run the DEVIATION program for the sample sizes listed below. Record the Sampling Distribution S.D. Mean and ERROR for each sample size.
   Note: ERROR = SAMPLING S.D. MEAN − POPULATION S.D.

N	SAMPLING S.D. Mean	POPULATION S.D.	ERROR
5	________________	28.58	____________
10	________________	28.58	____________
15	________________	28.58	____________
20	________________	28.58	____________
25	________________	28.58	____________
30	________________	28.58	____________
35	________________	28.58	____________
40	________________	28.58	____________
45	________________	28.58	____________
50	________________	28.58	____________

a. Does the SAMPLING S.D. MEAN consistently overestimate the POPULATION S.D.? YES _______ NO _______
b. Does the variance of the sampling distribution decrease with increased sample size? YES _______ NO _______
c. The mean of the sampling distribution doesn't get any closer to the population standard deviation. What do the signs of the differences lead you to think about the direction of the bias when the sample standard deviation is used as an estimator of the population standard deviation?

____

____

Confidence Intervals

Confidence intervals can be computed for many different population parameters by using the standard error of the statistic and the confidence level. A **standard error of a statistic** is computed for each type of sample statistic, e.g., standard error of the mean. The variance of a sampling distribution indicates the amount of error in estimating the population parameter. Smaller sampling variance reflects less error in estimating the population parameter. The standard error of a statistic is computed as the standard deviation of the sampling distribution divided by the square root of the sample size. Consequently, as sample size increases, the standard error of the statistic decreases. A **confidence interval** is computed using the sample statistic and the standard error of the statistic (standard deviation of the statistic in the sampling distribution). The confidence interval around the sample statistic is a range of values that should contain the population parameter. A **confidence level** is used which defines how confident we are that the interval around the sample statistic contains the parameter being estimated. The confidence interval is determined by picking an area in the tail of the sampling distribution in which the value of the statistic is improbable. Recall that a sampling distribution is a frequency distribution; therefore we could pick a 5% probability area, leaving 95% of the sample statistics in the frequency distribution as plausible estimates of the population parameter.

The confidence interval around the sample statistic is computed by using the standard error of the statistic. The **confidence interval** indicates the precision of the sample statistic as an estimate of the population parameter. The confidence interval around the sample statistic is said to include the population parameter with a certain level of confidence. It should be common practice to report confidence intervals for various population parameters such as proportions, means, or correlations. The confidence interval contains a high and low score, above and below the sample statistic, in which we feel confident that the interval contains the population parameter. **Confidence levels** are used to determine the confidence interval width. Commonly used confidence levels are 90%, 95%, or 99%. These confidence levels for sample data are indicated by a critical t-value of 1.65 (10% probability area), 1.96 (5% probability area), and 2.58 (1% probability area), respectively, which are given in Table 2 (Distribution of t for Given Probability Levels).

The 95% confidence interval for the population mean is computed as:

$$\bar{X} \pm 1.96(S / \sqrt{n})$$

where S = sample standard deviation and n = sample size. The value of 1.96 corresponds to the t-value that contains 5% of the sample means in the tail of the sampling distribution that are improbable. This implies that 5 times out of 100 replications, a confidence interval for a sample mean may not contain the population mean. In contrast, 95 times out of 100, the confidence interval around the sample mean will contain the population mean. Stated differently, 5% of the time the sample mean will not be a good estimate of the population mean, or conversely, 95% of the time we are confident that the sample mean will be a good estimate of the population mean.

If the sample mean was 50, sample standard deviation 10, the sample size 100, and we wanted to be 90% confident that the confidence interval captured the population mean, then the confidence interval around the sample mean would range between 51.65 and 48.35. This range of values is computed as:

$$CI_{.90} = \bar{X} \pm t(\frac{S}{\sqrt{n}}) = 50 \pm 1.65(\frac{10}{\sqrt{100}}) = 50 \pm 1.65$$
$$CI_{.90} = (51.65, 48.35)$$

If we replicated our sampling ten times, we would expect the population mean to fall in the range of values approximately 90% of the time (9 times out of 10 the confidence intervals would contain the population mean).

The 95% confidence interval using the population standard deviation would be computed as:

$$\bar{X} \pm 1.96(\sigma / \sqrt{N})$$

If the population standard deviation is known, one would use the population standard deviation with a z-value for the confidence interval. If the population standard deviation is *not* known, one would use a sample standard deviation estimate with a critical t-value for the confidence interval.

A confidence interval reflects a range of values (high, low) around the sample mean for different levels of confidence, e.g., 90%, 95%, and 99%. A confidence interval indicates the precision of a sample statistic as an estimate of a population parameter. If a confidence interval has a 95% level of confidence, this indicates that approximately 95 out of 100 confidence intervals around the sample statistic will contain the population parameter. If the confidence level remains the same, but the sample size increases, then the width of the confidence interval decreases, indicating a more precise estimate of the population parameter.

CONFIDENCE R Program

The CONFIDENCE program will simulate random samples and compute the confidence intervals around the sample mean (the process is the same for other population parameters because it would be based on the sampling distribution of the statistic). A population with a normal distribution ($\mu = 50$ and $\sigma = 10$) will be sampled. You can enter different sample sizes and confidence levels to see the effect they have on estimating the confidence interval. The program uses the population standard deviation rather than the sample standard deviation in the confidence interval formula because it is known. The sampling distribution of the mean will be based on 20 replications. For each sample mean, the 95% confidence interval around the mean will be computed and the program will check to see whether or not the population mean of 50 is contained in the confidence interval. Due to sampling error, one may not achieve the exact percent of confidence intervals that contain the population mean as indicated by the confidence level, i.e., 95%.

The program creates confidence intervals around repeated samples taken from a population and tests to see whether they contain the population mean. The sample size, population mean, population standard deviation, number of replications and size of the confidence interval can be changed in the program. The confidence interval size is specified as a z-value. Samples are simulated for the number of desired replications. The mean of each sample as well as the confidence intervals are calculated according to the formula given in the chapter. There is a count of the number of times the population mean is captured by the confidence interval for the sample. This is expressed as a percentage based on all the replications at the end of the program, both as a ratio and percent. Individual sample information is output, including the sample means and confidence intervals, which capture the population mean.

CONFIDENCE Program Output

```
Pop. Mean = 50 Pop. SD = 10
Sample Size = 100 N Replications = 20

Confidence Intervals for Z value = 1.96
Confidence Intervals that Contain Population Mean = 18 / 20 = 90 %

Sample Mean CI (high - low)   Pop.   Mean Within CI
48.16       50.12 - 46.2      50          Yes
49.58       51.54 - 47.62     50          Yes
49.42       51.38 - 47.6      50          Yes
49.92       51.88 - 47.96     50          Yes
50.58       52.54 - 48.62     50          Yes
52.78       54.74 - 50.82     50          No
49.34       51.3 - 47.38      50          Yes
50.44       52.4 - 48.48      50          Yes
50.09       52.05 - 48.13     50          Yes
49.62       51.58 - 47.66     50          Yes
49.7        51.66 - 47.74     50          Yes
```

51.05	53.01 - 49.09	50	Yes
50.59	52.55 - 48.63	50	Yes
49.85	51.81 - 47.89	50	Yes
51.29	53.25 - 49.33	50	Yes
50.47	52.43 - 48.51	50	Yes
47.2	49.16 - 45.24	50	No
50.73	52.69 - 48.77	50	Yes
50.04	52 - 48.08	50	Yes
49.7	51.66 - 47.74	50	Yes

Confidence Interval Exercises

1. Run the CONFIDENCE program for ten (10) replications with a confidence level of 90% (z = 1.65); keep the population mean of 50, standard deviation of 10, and sample sizes of 100.
   a. How many of the confidence intervals contained the population mean? ________________
   b. What percentage did you expect to contain the population mean? ________________
   c. If you increased the number of replications, would you more closely approximate the confidence level percent? Hint: Run CONFIDENCE and change the number of replications to observe what happens. YES _______ NO ________________
   d. Why does increasing the number of replications not guarantee a more close approximation to the confidence level percent? ______________ ________________

2. Run the CONFIDENCE program for fifty (50) replications with a confidence level of 90% (z = 1.65); keep the population mean of 50, standard deviation of 10, and sample sizes of 100.
   a. Do all the confidence intervals have the same width? YES ________ NO ________________
   b. What is the width of the confidence interval? ________________________
   c. Compute the confidence interval width using the population standard deviation and the sample size formula in the chapter. ______________________

$$\bar{X} \pm 1.65(\sigma / \sqrt{N})$$

   d. Does the formula give the same confidence interval width as the CONFIDENCE program? Note: (1.65 * 2 = 3.30; using the probability area in both tails of the sampling distribution). YES ________ NO ________

3. Run the CONFIDENCE program for each of the sample sizes listed below. Keep the population mean of 50, standard deviation of 10, and set the number of replications to 10 for a 1.96 confidence level (95%). Record the high and low values for the confidence intervals and calculate the width of the confidence interval.

Note: You can obtain the confidence interval width by subtracting the low value from the high value in the outputted table.

Confidence Interval for Z value = 1.96

Sample Size	CI (High-Low)	CI Width
10	____________	_______
144	____________	_______
256	____________	_______
625	____________	_______

a. As sample size increases, does the confidence interval width that contains the population mean become smaller? YES ______ NO ______
b. If the confidence interval width becomes smaller as sample size increases, does this imply a more accurate estimate of the population mean? YES ______ NO ______

4. Run the CONFIDENCE program for each of the confidence levels listed below. Keep the population mean of 50, standard deviation of 10, sample size of 100, and set the number of replications to 100. Record the CI high and low values and the percent of confidence intervals that contained the population mean.

Confidence Level	CI (High-Low)	Percent
90	____________	_______
95	____________	_______
99	____________	_______

a. Does the confidence interval become wider as the confidence level increases from 90% to 99%? YES _______ NO _______
b. If the confidence interval becomes wider, does this imply that we are more confident to have captured a range of values that contains the population mean. YES _______ NO _______

Statistical Hypothesis

The scientific community investigates phenomena in the world. The areas for scientific inquiry are many and have led to the creation of numerous academic disciplines, e.g., botany, biology, education, psychology, business, music, and so forth. The first step in any academic discipline that conducts scientific investigation is to ask a research question. Research questions can be expressed in many different ways. For example, “In the upcoming election, who will be elected President of the United States?” or “Which is better, margarine or butter, in lowering cholesterol?” The next important step is to design a study, then gather data and test the research question. This requires converting the research question into a statistical hypothesis.

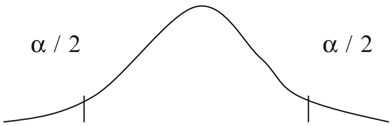
There are many different kinds of statistical hypotheses depending upon the level of measurement (nominal, ordinal, interval, or ratio) and type of research design used in the study. A statistical hypothesis is the cornerstone to testing the two possible outcomes, which are always stated in terms of population parameters, given the kind of data collected (percents, ranks, means, or correlation coefficients). The two possible outcomes of a statistical hypothesis are stated in a null (H_O: no difference) and alternative (H_A: difference exists) format using symbols for the population parameter. The alternative statistical hypothesis is stated to reflect the outcome expected in the research question. This involves either a directional (greater than) or non-directional (difference exists) expression. The null hypothesis is the corresponding opposite expected outcome of less than/equal or no difference, respectively. A research question and statistical hypothesis for each type of data is listed.

Research question	Data	Statistical hypothesis
Is the percent of people drinking beer in Texas greater than the national average?	Percents	H_O: $P_{Texas} \leq P_{National}$ H_A: $P_{Texas} > P_{National}$
Is there a difference in the ranking of persons on two different diets for weight gain?	Ranks	H_O: $R_{Diet\ A} = R_{Diet\ B}$ H_A: $R_{Diet\ A} \neq R_{Diet\ B}$
Does my 5th grade class on average score higher than the national average in math?	Means	H_O: $\mu_{Class} \leq \mu_{National}$ H_A: $\mu_{Class} > \mu_{National}$
Is the relationship between music ability and self-esteem in my sample of students different than the population?	Correlation	H_O: $\rho_{Sample} = \rho_{Population}$ H_A: $\rho_{Sample} \neq \rho_{Population}$

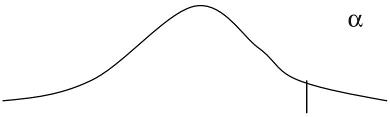
Two kinds of errors are associated with our decision to retain or reject the null hypothesis based on the outcome of the statistical test (TYPE I error and TYPE II error). The TYPE I error is specified by selecting a level of significance (probability area) such that if the sample statistic falls in this probability area, then we reject the null hypothesis in favor of our alternative hypothesis. The TYPE II error corresponds to the probability of retaining the null hypothesis when it is false. When we state the statistical hypothesis as "greater than," we designate only one-tail of the sampling distribution of the statistic because of the directional nature of the research question. When we state the statistical hypothesis as a "difference exists," we designate both tails of the sampling distribution of the statistic because of the non-directional nature of the research question. Consequently, the probability area corresponds to different "tabled statistics."

Once the probability area is determined for the statistical hypothesis, we can select a tabled statistical value to set our region of rejection. The tabled statistical values were generated using probability theory and created from the sampling distribution of the statistic for different sample sizes (degrees of freedom) and levels of significance. Only the more popular levels of significance (.05, .01, and sometimes .001) are included in the tables due to page length consideration. Consequently, it is common for researchers to select a region of rejection and test statistical hypotheses based on the .05, .01, or .001 levels of significance. The relationship between the level of significance and probability area for the region of rejection (vertical hash marks) can be depicted as follows:

Non-directional (two-tailed) Research Question:



Directional (one-tail) Research Question:

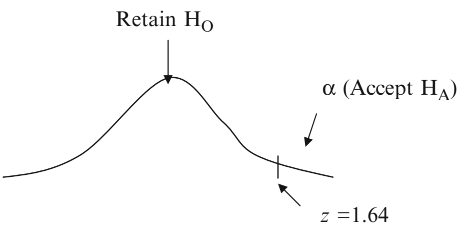


An example research question and corresponding statistical hypothesis will help to illustrate the relationship between the non-direction and/or direction of the question, level of significance, and region of rejection. The research question, "Is the SAT population mean in Texas greater than the SAT population mean for the U.S.," is converted into a null and alternative statistical hypothesis:

$$H_O : \mu_{\text{Texas}} \leq \mu_{\text{U.S.}}$$

$$H_A : \mu_{\text{Texas}} > \mu_{\text{U.S.}}$$

This is a directional research question, hence the alternative statistical hypothesis indicates a greater than expression, while the null statistical hypothesis indicates less than or equal to for the population parameters. We test our hypothesis by random sampling of 100 students in Texas, compute the sample mean, and conduct a statistical test at the .05 level of significance. Once we have selected a sample size and level of significance for the study, a tabled statistical value can be selected. Under the normal probability curve in Table 1, a z-value of 1.64 corresponds to a probability value (p-value) that indicates an area approximately equal to .05 (probability area beyond z). This z-value is used to set the region of rejection for testing our statistical hypothesis: $R_{.05} = z > 1.64$. When we conduct a z-test for the difference between the means, a computed z-value greater than 1.64 will imply that the population mean in Texas is greater than the population mean in the U.S. In other words, the computed z-value falls in the probability area of the sampling distribution of the statistic that we have designated to be a highly improbable outcome. The probability area for the null hypothesis and the alternative hypothesis is therefore depicted along with the tabled z-value and level of significance for a directional research question as follows:



Given our selection of the .05 level of significance, only 5% of the time would we expect a mean difference to exceed $z = 1.64$ if the null hypothesis is true, i.e., the population means are equal. Consequently, if the mean difference is large enough and we compute a z-value greater than 1.64, we are 95% confident in our decision to reject the null hypothesis in favor of the alternative hypothesis.

Assume we computed a Texas SAT mean of 530 for the 100 randomly sampled students and the U.S. SAT mean was 500 with a population standard deviation of 100. A one-sample z-test to determine the statistical significance of this mean difference is computed as:

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{N}} = \frac{530 - 500}{100 / \sqrt{100}} = \frac{30}{10} = 3.0$$

Since the computed $z = 3.0$ is greater than $z = 1.64$, and therefore falls in the region of rejection, we reject the null hypothesis of no difference in favor of the alternative hypothesis that the population SAT mean in Texas is greater than the population SAT mean in the U.S. We are 95% confident in our decision, but also know that 5% of the time our decision might be wrong (TYPE I error). We would answer the research question by stating that the population SAT mean in Texas is statistically significantly higher than the population SAT mean in the U.S.

This example used a z-test because the population standard deviation for the SAT was known. In many research studies, the population standard deviation is unknown, so we would use a t-test. The t-test formula that uses the *sample standard deviation* in place of the population standard deviation is:

$$t = \frac{\bar{X} - \mu}{S / \sqrt{N}}$$

Research questions involving a test of differences in population means are commonplace in several academic disciplines. An engineer needs to know the average weight of vehicles that can safely travel across a bridge. A psychologist needs to test whether a group of clients have an average cognitive ability greater than the national average. A sociologist needs to determine the average family size for different ethnic groups. The auditor of a large oil company wants to know the average amount of error in the bills prepared for customers. A doctor studying the possible reduction in blood pressure caused by a new medicine is concerned about the average reduction in blood pressure for patients. These and many other research questions reflect a need for a statistical procedure to test whether population means are statistically different.

Tests of statistical hypotheses do not provide exact outcomes. Instead, the tests are based on data-gathering evidence to reach a conclusion with some degree of uncertainty. In other words, it is possible to reject the null hypothesis when in fact the null hypothesis is true. We preset the probability of making this kind of error (TYPE I error) when selecting the level of significance and corresponding tabled statistic, which is based on the sampling distribution of the statistic. We have already learned that increasing sample size, using a directional hypothesis, willingness to detect a larger difference, and the level of significance are factors that influence whether the statistical test is powerful enough to detect a difference when in fact one exists.

In testing whether the population SAT mean in Texas was significantly different from the population SAT mean in the U.S., knowledge of the population standard deviation and sample size played a role. A larger random sample of students would dramatically reduce the standard error in the formula, which would result in a larger computed z-value, e.g., $z = 30$ if $N = 10{,}000$. A smaller sample size would result in a smaller z-value. If the population standard deviation is *not* known, a sample standard deviation as an estimate might produce a very different result.

In hypothesis testing, the null hypothesis is retained as true unless the research findings are beyond chance probability (unlikely to have occurred by chance). A TYPE I error occurs when a true null hypothesis is rejected erroneously, usually due to an atypical research outcome. The region of rejection is specified by the level of significance (α), which indicates the probability of making a TYPE I error. If the z-value falls in the region of rejection probability area, then the null hypothesis is rejected in favor of the alternative hypothesis. For different values of alpha (levels of significance) and sample size, the region of rejection will be indicated by a tabled statistic from the sampling distribution of the statistic (Appendix). A TYPE II error occurs when a false null hypothesis is retained erroneously, usually due to insufficient data.

HYPOTHESIS TEST R Program

The HYPOTHESIS TEST program allows you to specify the true population mean and then test various null and alternative hypotheses. In the program output, you will be able to observe either z-tests or t-tests for several statistical hypotheses depending upon whether the population standard deviation (z-test) or sample standard deviation (t-test) is used. The program will select a random sample of size N from the true population, compute the sample mean and variance, compute the appropriate test statistic, the p-value, and indicate the decision outcome. Since you specify the true population mean, you will know whether or not the decision to reject the null hypothesis is correct. If the population standard deviation is used (set varUse = 0, the default), the z-statistic is reported. If the sample standard deviation is used (set varUse = 1), the t-statistic is reported.

The program uses the **pValue** function in R to determine the probability in the tails for either the z-test or t-test after all the user-defined variables are initialized. Based on the direction of the statistical hypothesis for the z-test or t-test, probability values are determined using the **pnorm** function, which returns the probability value for a mean difference with a given standard deviation. If the sample mean is less than the null hypothesis mean, then **pnorm** is reported; if the sample mean is greater than the null hypothesis mean, then **1-pnorm** is reported. This use of the

pnorm function results in only the probability area for one-tail of either the normal- or t-distribution. For a two-tailed test, the probability area (alpha) is divided evenly between the two ends of the distributions. The program specifies the probability area for one- and two-tailed tests by selecting a value for the **tails** variable. The program default is $p<.05$, therefore if the computed p-value is less than .05, the decision is reject the null. If the p-value is greater than .05, the decision is retain the null. The number of statistical tests computed and printed is based upon the value for **numSamples**.

HYPOTHESIS TEST Program Output

```
z Statistic

Pop. Mean=10.5 Pop. Variance=2 Null Mean=10
Sample Size=36 Alpha=0.05 Number of Samples 10

Variance type=0 (0=population; 1=sample)
Hypothesis direction=1 (0<Null, 1>Null, 2=two-tailed)

Sample Mean   Pop.  SD z-statistic   Decision      p-value
10.251        1.414    1.064         RETAIN NULL   0.144
10.855        1.414    3.626         REJECT NULL   0.001
10.521        1.414    2.209         REJECT NULL   0.014
10.262        1.414    1.113         RETAIN NULL   0.133
10.265        1.414    1.126         RETAIN NULL   0.131
10.65         1.414    2.756         REJECT NULL   0.003
10.704        1.414    2.985         REJECT NULL   0.002
10.719        1.414    3.049         REJECT NULL   0.002
10.629        1.414    2.668         REJECT NULL   0.004
10.653        1.414    2.771         REJECT NULL   0.003

t statistic

Pop. Mean=10.5 Pop. Variance=2 Null Mean=10
Sample Size=36 Alpha=0.05 Number of Samples 10

Variance type=1 (0=population; 1=sample)
Hypothesis direction=1 (0<Null, 1>Null, 2=two-tailed)

Sample Mean  Sample SD  t-statistic    Decision       p-value
10.429       1.226      2.097          REJECT NULL    0.018
10.678       1.561      2.607          REJECT NULL    0.005
10.784       1.568      2.998          REJECT NULL    0.002
10.681       1.469      2.783          REJECT NULL    0.003
10.47        1.446      1.95           REJECT NULL    0.026
10.036       1.423      0.151          RETAIN NULL    0.441
10.021       1.817      0.07           RETAIN NULL    0.473
9.919        0.933      -0.521         RETAIN NULL    0.699
10.338       1.416      1.433          RETAIN NULL    0.076
10.309       1.301      1.423          RETAIN NULL    0.078
```

Hypothesis Testing Exercises

1. Run the HYPOTHESIS TEST program with the following initial values which defines a one-tail test and uses the population variance:

```
popMean <- 10
popVar <- 2
nullMean <- 10
tails <- 1
sampleSize <- 36
alpha <- .05
varUse <- 0
numSamples <- 10
```

Sample	Sample Mean	z-statistic	Decision	p-value
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

a. Are the z-statistics correct in the table (Use the formula below to verify)?
YES ________ NO ________ $z = \frac{\overline{X} - \mu}{\sigma / \sqrt{N}}$

b. Compare each p-value to $\alpha = 0.05$. If $p \geq 0.05$, then the decision should be to RETAIN the null hypothesis. If $p < .05$, then the decision should be to REJECT the null hypothesis. Do all of the decisions correspond to the p-values?
YES ________ NO ________

c. Since the null hypothesis is actually true, any REJECT NULL decisions are errors. This is a TYPE I error. The probability of a TYPE I error is equal to α. How many REJECT NULL decisions were there? What percent is this?
____

2. Run HYPOTHESIS TEST program again using all of the same values, except increase the number of samples to 40 (numSamples <- 40).

a. What is the percent of TYPE I errors in the 40 simulations? __________
b. What do you expect this TYPE I error percent to be? __________

3. Run HYPOTHESIS TEST program again, except change alpha to .10 (alpha <−.10).

a. How many times was the null hypothesis rejected? __________
b. What do you expect this percent to be? __________

4. Run HYPOTHESIS TEST program again, except increase the sample size to 100 (sampleSize<−100).

   a. How many times is the null hypothesis rejected? _________
   b. Does sample size have an effect on the percent of TYPE I errors? _________

5. Run HYPOTHESIS TEST program again, except use the sample variance (varUse<−1). In most cases the proportion of rejections will be much higher because the sample standard deviation is not as good an estimator as the true population standard deviation.

   a. What is the percent of rejections? _________
   b. Does the percent of rejections approximate α? YES _____NO ______

6. Run the HYPOTHESIS TEST program in which the null hypothesis is false (The population mean is actually greater than the null mean). Run the program five times using a different sample size each time. Use the following settings:

```
popMean <- 10.5
popVar <- 2
nullMean <- 10
tails <- 1
sampleSize <- 36 (Use different sample sizes: 25, 16, 8, 2)
alpha <- .05
varUse <- 0
numSamples <- 20
```

   Record the number of times the null hypothesis was retained or rejected for each run. Compute the percentage of null hypotheses that were retained. This is the probability of a TYPE II error.

Sample Size	Number Rejected	Number Retained	Percent Retained
36			
25			
16			
8			
2			

   a. Which sample size had the greatest probability of a TYPE II error?

      36 ____ 25 ____ 16 ____ 8 _____ 2 _____
   b. What effect does sample size have on the probability of a TYPE II error? Hint: A TYPE II error is when you retain a false null hypothesis. ___________

7. Run the HYPOTHESIS TEST program again in which the null hypothesis is false. Use alpha<- .10. Run the program five times using the same sample sizes as before. The program settings should be:

```
popMean <- 10.5
popVar <- 2
nullMean <- 10
```

```
tails <- 1
sampleSize<- 36 (Use different sample sizes: 25, 16, 8, 2)
alpha <- .10
varUse <- 0
numSamples <- 20
```

Sample Size	Number Rejected	Number Retained	Percent Retained
36			
25			
16			
8			
2			

a. Which sample size had the greatest percent retained (probability of a TYPE II error)? 36 ____ 25 ____ 16 ____ 8 _____ 2 _____
b. Did an increase in alpha from .05 to .10 *decrease* the percent retained (probability of a TYPE II error)? YES _______ NO _______
c. Did an increase in alpha from .05 to .10 *increase* the percent rejected (probability of a TYPE I error)? YES _______ NO _______

8. Run the HYPOTHESIS program again, except this time use a two-tailed test. Run the program five times using the same sample sizes as before. The program settings should be:

```
popMean<- 10.5
popVar<- 2
nullMean<- 10
tails<- 2
sampleSize<- 36 (Use different sample sizes: 25, 16, 8, 2)
alpha<- .10
varUse<- 0
numSamples<- 20
```

Sample Size	Number Rejected	Number Retained	Percent Retained
36			
25			
16			
8			
2			

a. Does a two-tailed test, in comparison to the previous one-tail test, result in more Null Hypotheses being retained? YES ______ NO ______
b. Does sample size affect the results? YES ______ NO ______

TYPE I Error

Science, as a way of understanding the world around us, has for centuries encompassed the classification, ordering, and measuring of plants, characteristics of the earth, animals, and humans. Humans by their very nature have attempted to understand, explain, and predict phenomena in the world around them. When individuals, objects, or events are described as a well-defined, infinite population, random sampling and probability statistics can play a role in drawing conclusions about the population. The problem in Science is essentially one of drawing conclusions about the characteristics of infinitely large populations.

The use of a random sample to compute a statistic as an estimate of a corresponding population parameter involves some degree of uncertainty. For example, the sample mean doesn't always fall close to the population parameter (sampling error) and therefore isn't always accurate in the estimation of the population parameter. However, we can use the number of times the sample mean falls in an area under the sampling distribution to indicate a degree of confidence. Recall, the areas were designated as 68%, 95% or even 99%. These areas are called the confidence level (probability of occurrence). The number of times the sample mean falls outside these areas (confidence interval) we refer to the result as committing a TYPE I Error (probability of non-occurrence). The **TYPE I Error** therefore indicates the amount of uncertainty or probability of error, especially when making a decision about a population parameter. We generally make a decision about a population parameter in the context of a research question.

The research question could be whether or not to use a new experimental drug for the treatment of a disease. In other words, how confident are we that the drug will work. It could also be whether or not to spend money on an innovative math program for High School students. How confident are we that the innovative math program will be better than the traditional math program? In business, the research question of interest could be whether or not to implement a new strategic plan. We conduct research, formally or informally, to answer these types of questions. In simple terms, we ask a question, gather data, and then answer the question. This is the essence of the research process. However, in making our decision to release a drug for public use or spend thousands of dollars on an innovative math program, we can never be 100% certain it will work.

The research process involves the formulation of a question that can be tested, the collection of relevant data, the analysis and presentation of the data, and the answering of the question. This formal research process embodies the use of the following scientific principles:

a. Statement of the Research Hypothesis
b. Selection of Sample Size and Sample Statistic
c. Selection of Confidence Level and Region of Rejection
d. Collection and Analysis of Data
e. Statistical test and Interpretation of Findings

In conducting research using random samples from a population, our research hypothesis is related to the probability of whether an event occurs or not. The probability of an event occurring and the probability of an event not occurring are equal to 100% (see Chap. 2). As researchers, we accept some level of probability or uncertainty as to whether an event occurs or not. Our statistical test and interpretation of findings are linked to the TYPE I Error in our decision?

The statement of the research hypothesis expresses the outcomes expected in the research question. The outcomes are expressed as a Null Hypothesis and an Alternative Hypothesis. The **Null Hypothesis** is a statement of no difference between the sample mean and population parameter, or in general terms between a sample statistic and a population parameter. The **Alternative Hypothesis** is a statement that a difference exists between the sample mean and population parameter, typically based upon some intervention, treatment, or manipulation of participants, objects, or events. For example, the Null Hypothesis would state *no difference* in the average mathematics test scores of students in a traditional math program versus an innovative math program. The Alternative Hypothesis would state that the average mathematics test score of students in the innovative math program is greater (*statistically different*) than the average mathematics test score of students in the traditional math program.

The sample size and sample statistic chosen to make a comparison between the sample mean and population mean is determined next in the research process. The sample size is an important consideration because as sample size increases the sample statistic more closely approximates the population parameter. The sample statistic is computed by taking the difference between the sample mean and the population mean, or between two sample means, divided by the standard error of the mean difference, which is based on the sampling distribution of the statistic. We will learn more about conducting these types of statistical tests in later chapters.

The confidence level and region of rejection are now established which set the amount of uncertainty we are willing to accept for the two possible outcomes of the research question, i.e., null versus alternative. If we want to be 95% confident in our decision that the innovative math program produced higher average mathematics test scores, then we must also have a 5% level of uncertainty (TYPE I Error) in our decision. This probability of making a mistake (uncertainty) is called the **level of significance** and denoted by the symbol, α. It is the chance we take of rejecting the Null Hypothesis statement when it is true and erroneously accepting the Alternative Hypothesis statement. If we *reject* the Null Hypothesis statement when it is true, and *accept* the Alternative Hypothesis statement, we commit a **TYPE I Error**. If we *retain* the Null Hypothesis statement when it is false, thus *don't accept* the Alternative Hypothesis statement, we commit a **TYPE II Error**, which will be discussed in the next section. In either instance, we do so with a level of confidence and a level of error in our decision.

The region of rejection refers to the selection of a statistical value from the sampling distribution that is at the cut-off point for the beginning of the probability area in which we would reject the null hypothesis in favor of the alternative hypothesis. Statistical values for varying sample sizes and different types of sampling distributions for the 90%, 95%, and 99% confidence levels can be found in the appendix of most statistics books. We refer to these statistical values as the "tabled statistic." If the sample statistic computed from the sample data falls in the 5% area, corresponding to a 95% confidence level, then we reject the Null Hypothesis and accept the Alternative Hypothesis. We make this decision knowing that 5% of the time it might not be the correct decision; which is the TYPE I Error.

The research process would now involve randomly sampling students and randomly assigning them to either a traditional math program or an innovative math program. After a semester of study, each group would take the same mathematics test. The sample means and standard deviations for each group would be computed. A statistical test would determine if the means were statistically different for our level of confidence and corresponding level of significance. In this instance, an independent t-test would be computed which will be presented in a later chapter (see Chap. 10).

The final step in the research process is a comparison of the "tabled statistic" (based on the level of significance and region of rejection) and the sample statistic computed from the sample data. If the sample statistic is greater than the tabled statistic, a decision is made to reject the Null Hypothesis and accept the Alternative Hypothesis. If the sample statistic does not exceed the tabled statistic, then a decision is made to retain the Null Hypothesis and reject the Alternative Hypothesis. Our interpretation of the findings of the research process is based on this decision to reject or accept the research question, which relates to the two possible outcomes of the study.

TYPE I Error is the probability of rejecting a Null Hypothesis statement, which is true, and erroneously accepting an Alternative Hypothesis. TYPE I Error probability is the same as the level of significance denoted by the symbol, α. As the level of confidence is increased, TYPE I Error is decreased, and the width of the confidence interval increases. As the sample size is increased, the width of the confidence interval is decreased. The sample mean is always in the confidence interval. The population mean is sometimes outside of the confidence interval; the percentage of the intervals that contain the population mean is the confidence level, while the percentage of the intervals that don't contain the population mean is the TYPE I error probability.

TYPE I ERROR R Program

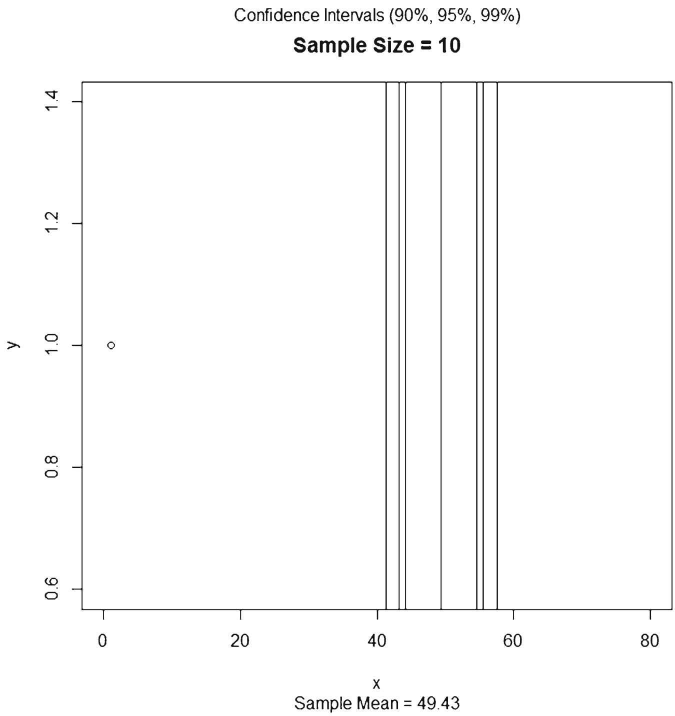
The TYPE I ERROR program shows the relationship between the confidence level and the width of the confidence interval, but in the context of TYPE I Error. In addition, you can use different sample sizes to observe the effect of sample size on the width of the confidence interval, the confidence level, and the TYPE I Error. The program is initially set with a population mean of 50 and a standard deviation of 10. The TYPE I ERROR program computes and graph the 90%, 95%, and 99% confidence intervals which correspond to the 10%, 5%, and 1% levels of uncertainty (levels of significance). Since the population standard deviation is known, z-values rather than t-values are used to compute the intervals. The tabled z-values that correspond to the levels of significance are 1.65, 1.96, and 2.58, respectively, for 90%, 95%, and 99% confidence levels. The program prints the location of the sample mean so that you can compare it with the true population mean of 50. The vertical lines, starting with the two inside lines, indicate the 90%, 95%, and 99% confidence intervals, respectively.

TYPE I ERROR Program Output

```
Pop. Mean=50 Pop. SD=10

Sample Size=10 N Replications=100

Confidence Level  Percent Error  Confidence Interval   Interval
                                                      Width
90%               10%            54.6 - 44.2           10.4
95%               4%             55.6 - 43.2           12.4
99%               3%             57.6 - 41.3           16.3
```



TYPE I Error Exercises

1. Run the TYPE I ERROR program for a sample size of 100; keep the other values the same. Record the confidence interval, interval width and TYPE I Error.

Confidence Level	Confidence Interval	Interval Width	TYPE I Error
90%	________________	____________	__________
95%	________________	____________	__________
99%	________________	____________	__________

a. Does the Interval Width become larger as the Confidence Level increases from 90% to 99%? YES ______ NO ______
b. Does the TYPE I Error become smaller as the Confidence Level increases from 90% to 99%? YES ______ NO ______
c. Which confidence level has the greatest probability of *not* containing the population mean? 90% _____ 95% _____ 99% _____
d. Were the TYPE I Error percents the same as the Expected Error percents?

TYPE I Error	Expected Error	Confidence Level
_________	10%	90%
_________	5%	95%
_________	1%	99%

2. Run the TYPE I ERROR Program for the sample sizes listed below. Record the Confidence Interval Width and TYPE I Error for each sample size and confidence level.

90% Confidence Level

Sample Size	Interval Width	TYPE I Error
10	__________	_________
100	__________	_________
500	__________	_________
1000	__________	_________

95% Confidence Level

Sample Size	Interval Width	TYPE I Error
10	__________	_________
100	__________	_________
500	__________	_________
1000	__________	_________

99% Confidence Level

Sample Size	Interval Width	TYPE I Error
10	__________	_________
100	__________	_________
500	__________	_________
1000	__________	_________

a. Does the confidence interval width become smaller as sample size increases for 90%, 95%, and 99% confidence levels. YES ______ NO ______
b. In comparing the results for large versus small sample sizes across the three confidence levels, do the TYPE I errors become closer to the level of confidence as the sample size increases? YES ______ NO ______
c. Is it possible for the sample mean to be outside of the confidence interval? YES ______ NO ______

TYPE II Error

The research process begins with the formulation of a research question. The research question is then stated in a form that can be tested which is called a statistical hypothesis. The two possible outcomes of the statistical hypothesis are termed a **null hypothesis** and an **alternative hypothesis**. The null hypothesis is stated as "no difference" between population parameters and indicates, "to be nullified." The alternative hypothesis reflects the outcome expected by the research question. Consequently, the alternative hypothesis can be stated in either a directional or non-directional format. A **directional** hypothesis states that one population parameter is greater than the other population parameter. A **non-directional** hypothesis states that the two population parameters "are different," but doesn't specify which is greater than the other population parameter. In the previous chapter we assumed a non-directional hypothesis because we used the probability area in both tails (both sides) of the normal bell-shaped probability curve. It is also important to point out that statistical hypotheses are stated in the form of population parameters.

An example of both the non-directional and directional statistical hypotheses will help to visualize how they are presented using population parameters. The **non-directional** hypothesis in both Null Hypothesis (H_O) and Alternative Hypothesis (H_A) form, stating that Company A and Company B have the same average sales, would be indicated as:

$$H_O : \mu_{CompanyA} = \mu_{CompanyB}$$
$$H_A : \mu_{CompanyA} \neq \mu_{CompanyB}$$

The Null Hypothesis indicates that the population means for the two companies are equal. The Alternative Hypothesis indicates that the population means for the two companies are different.

The **directional** hypothesis in both Null Hypothesis (H_O) and Alternative Hypothesis (H_A) form, stating that Company A had greater average sales than Company B, would be indicated as:

$$H_O : \mu_{CompanyA} \leq \mu_{CompanyB}$$
$$H_A : \mu_{CompanyA} > \mu_{CompanyB}$$

The Null Hypothesis indicates that the population mean for Company A is less than or equal to the population mean for Company B. The Alternative Hypothesis indicates that the population mean for Company A is greater than the population mean for Company B, thereby only testing a specific directional difference in the population means of the companies.

The probabilities related to the two possible outcomes, Null Hypothesis and Alternative Hypothesis, are given the name TYPE I error and TYPE II error.

A **TYPE I error** is when the null hypothesis is rejected, but in fact it is true. This means that you gathered evidence (data) and concluded that the evidence was strong enough to accept an alternative hypothesis, but did so erroneously. A **TYPE II error** is when the null hypothesis is accepted, but in fact it is false. This means that you gathered evidence (data) and concluded that the evidence wasn't strong enough to reject the null hypothesis, so you retained the null hypothesis, but did so erroneously. Not rejecting the null hypothesis was due to a lack of sufficient evidence (data). Neither decision, to reject or retain the null hypothesis, is 100% certain; therefore we make our decision with some amount of uncertainty. Whether the TYPE I error or TYPE II error is more important depends on the research situation and the type of decision made. An example will help to clarify the TYPE I and TYPE II errors and possible outcomes.

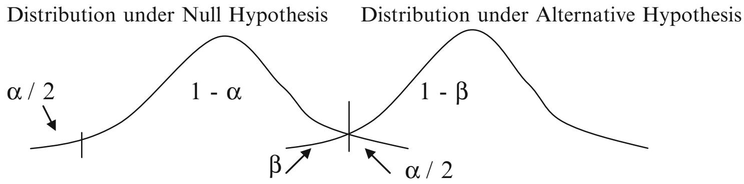
A corporation is going to use data from a sample of people to test a research question of whether or not a new product design will increase the average sales above that of the current product design. Using the current product design, the company sold an average of 85 cars. The company, in planning the study, realizes that two types of error are possible in the study. If the new product design is used, the average sales of the product may not be greater than 85 cars; but, the analysis could incorrectly lead to the conclusion that the average sales were higher, a TYPE I error. If the new product design was used and the average sales did exceed 85 cars, but the statistical analysis failed to indicate a statistically significant increase, then a TYPE II error will occur.

When planning this study, the company can control the probability of a TYPE I error by setting the level of significance (α) to a stringent probability level. For example, setting the level of significance to .05 implies that 95 times out of 100 replications of the study, a correct decision would be made, and 5 times out of 100 an incorrect decision would be made. Setting the level of significance to .01 implies that 99 times out of 100 a correct decision would be made, and only 1 time out of 100 an incorrect decision would be made. These probabilities for a correct decision and an incorrect decision are applicable only when the Null Hypothesis is true. Therefore, we set the TYPE I error to a probability level that instills confidence in our decision to reject the null hypothesis (no difference) and accept an alternative hypothesis (difference exists). A TYPE II error requires a further concern for many other factors in the study. The probability of a TYPE II error depends on the level of significance (B), the direction or non-direction of the research question, the sample size, the population variance, and the difference between the population parameters we want to be able to detect (effect size). These concerns determine how powerful our test will be to detect a difference when it actually exists in the population(s) we study. In planning the study, the company needs to be concerned about both TYPE I and TYPE II errors, as well as the consequences of the decision made based on the outcome of the decision.

In testing hypotheses, four possible outcomes exist with regard to the decision made. They are:

		Actual Population Condition	
		Null Hypothesis is True	Null Hypothesis is False
Decision	Reject Null Hypothesis	TYPE I ERROR (Probability = α)	CORRECT DECISION (Probability = $1 - \beta$)
	Retain Null Hypothesis	CORRECT DECISION (Probability = $1 - \alpha$)	TYPE II ERROR (Probability = β)

The probability of making a TYPE I error in the Null Hypothesis is denoted by the symbol alpha, α, and referred to as the level of significance, with the probability of a correct decision $1-\alpha$. The probability of making a TYPE II error in the Alternative Hypothesis is denoted by the symbol Beta, β, with the power of the test given by $1-\beta$; the probability of making a correct decision. The relationship between TYPE I error and TYPE II error is typically graphed using vertical hash marks to denote the area for the region of rejection. The two graphs below indicate the probability areas for the four decision outcomes.



The probability in the tails of the distributions can change because the null and alternative hypotheses can be stated in directional or non-directional forms. In this example, the hypotheses are in a non-directional form, i.e., Is there a statistically significant mean difference between boys and girls on the mathematics test? If the level of significance is .05, then one-half or .025 would be the probability area in the tails of the distribution. If we chose only the one tail ($p = .025$), then the alternative research question would be stated as directional, e.g., Will the girls on average score higher than the boys in mathematics achievement? The statistical tables in the appendix adjust for the directional versus non-directional nature when selecting a "tabled statistic" value.

The basic concern in selecting a TYPE II error is the value for power, which is $1-\beta$. **Power** is the probability of *not* committing the error. In conducting research in education, psychology, business, and other disciplines, we are concerned about the power of our statistical test to detect a difference in the population parameters. A graph of the power values against different population parameter values will result in *power curves* for different levels of significance and sample size. The resulting power curves provide an adequate determination of the power of a test for all alternative values of the population parameter.

The concerns in research for having an adequate sample size, the nature of TYPE I and TYPE II errors, power of a test, and selection of alternative values for the population parameters, called effect size, form the basis of hypothesis testing. The hypothesis testing theory, known as the *Neyman-Pearson* hypothesis-testing theory, formulates the relationship between these concerns. In practice, we might select power equal to .80, use a .05 level of significance, and a directional hypothesis to test our statistical hypothesis. The **effect size** or magnitude of difference between the population parameters also affects the power of the test. If a small difference must be detected, then the sample size, level of significance, and power must be increased. We also generally know that the power of a test increases as sample size increases, and the power of a test increases as the level of significance is decreased, i.e., .01 to .05. Basically, all of these values are inter-related such that increasing or decreasing one affects the others. There are software programs available on the Internet that help in selecting sample size or power for different statistical tests, given values for the other criteria, e.g. GPower 3.1.

A TYPE II error implies that a null hypothesis has been accepted, but is false. The probability of a TYPE II error is denoted by the symbol B. The probability of a TYPE II error decreases as α increases (e.g., .01 to .05), N increases, the population variance decreases, and the effect size is larger (true mean is farther from the hypothesized mean in the direction of the alternative hypothesis). When a TYPE II error decreases, power increases. Power is the probability of rejecting a null hypothesis that is false. When the null hypothesis is false, power is equal to $1-B$. When the null hypothesis is true, power is equal to α.

TYPE II ERROR R Program

The probability of a TYPE II error is affected by four criteria (alpha [α], sample size, population variance, and effect size) specified in the **TYPE II ERROR** program. The program inputs these values and determines whether or not to reject the null hypothesis. The program is initially repeated 100 times. The relative frequency of occurrence with which a false null hypothesis is retained is an approximation of the probability of a TYPE II error, B. The power of the test is calculated as 1 minus this relative frequency. Power is equal to $1-B$ when the null hypothesis is false. The theoretical values for the probability of retaining the null hypothesis and power are also given. If the true mean is equal to the mean in the null hypothesis, then the probability of retaining the null hypothesis is equal to $1-\alpha$. In this case, power is equal to α, the level of significance. The hypothesis testing in the program is directional to determine whether the mean of a population with a normal distribution is greater than a specified value, therefore, the level of significance is in one tail of the distribution.

In the program, samples are drawn for a given number of replications instead of for a single sample. A **p-value function** determines the probability area based on a directional one-tail test. The number of times the null hypothesis is rejected based on the p-value is summed. This sum is used along with the total number of replications to calculate the estimated probability of retaining the null hypothesis and estimated power.

Next, the critical value for the test is determined by using a new function called **qnorm**, which takes a probability, mean and variance and returns the number of standard deviations above or below the mean. This is multiplied by the standard deviation and added to the mean to obtain the raw score value that equates to a cutoff point for the region of rejection in the tail of the distribution. This cut-off value is used to obtain the true probability of retaining the null hypothesis and true power for the given number of replications. Two separate tables are output, one table for the null mean difference, population mean difference, population variance, sample size, and alpha; and one table for the estimated probability of retaining the null mean difference, estimated power, true probability of retaining the null mean difference, and true power.

TYPE II ERROR Program Output

```
Null Mean=0 Pop. Mean=0 Pop. Variance=1

Sample Size=5 Alpha=0.05 N Replications=100
Hypothesis Direction=1 (0=one-tailed<Null, 1=one-tailed>Null

Table 1. Means, Variance, Sample Size, and Alpha
Null Diff.   Pop Diff.   Pop Variance   Sample Size   Alpha
      0          0              1             5      0.05

Table 2. Estimated Null Mean Diff. % and Power with True Null %
and Power
Retain Null %   Power Estimate   True Null %   True Power
0.980           0.020            0.950         0.050
```

TYPE II Error Exercises

1. Run the TYPE II ERROR program for the values indicated in the following table for 100 replications.

Null Mean	Population Mean	Population Variance	Sample Size	Alpha	Retain Null %	Power Estimate	True Null %	True Power
0	0.0	1.0	10	.05				
0	1.0	1.0	10	.10				
0	1.0	1.0	10	.05				
0	1.0	1.0	10	.025				
0	1.0	1.0	10	.01				
0	1.2	1.0	10	.05				
0	1.5	1.0	10	.05				
0	2.0	1.0	10	.05				
0	1.0	1.5	10	.05				
0	1.0	2.0	10	.05				

a. In the first run, the Null Mean equals the Population Mean (Null Hypothesis is true), so does the True Power equal Alpha? YES ________ NO _________.
b. In runs 2–5, alpha decreases from .10 to .01 with other factors held the same, so does B increase (Retain Null %)? YES ________ NO _________.
c. In runs 2–5, alpha decreases from .10 to .01 with other factors held the same, so does Power decrease (Power Estimate)? YES ________ NO _________.
d. Does B (Accept Null %) and $1-B$ (Power Estimate) equal 100%? YES ________ NO _________.
e. In runs 6–8, the Population Mean is increasingly different from the Null Mean; that is, the population mean is getting farther away from the null mean in the direction of the alternative hypothesis (effect size). So with other factors held constant, does power increase? YES ________ NO _________.
f. In runs 9–10, the Population Variance is increased, so does Power decrease? YES ________ NO _________.
g. Try to summarize the relationships between alpha, B (Retain Null %), 1 - B (Power Estimate), and effect size in examples **a** to **f** above. ______________
____

2. Run TYPE II ERROR program with the following sample sizes: 10, 20, 30, and 40. Keep the other values the same in the program:

```
nullMean <- 0
popMean <- 1
popVar <- 1
sampleSize <- 10
alpha <- .05
tails <- 1
numReplications <- 100
```

a. As sample size increases with the other factors held constant, does B approach zero? YES ______ NO _______
b. As the sample size increases with the other factors held constant, does Power increase? YES ______ NO _______
c. What is the region of rejection for the directional hypothesis, or where the null hypothesis will be rejected, when n = 10? Note: Tabled Statistic = Mean + (z * Standard error of the statistic). ________________
____

3. Run TYPE II ERROR again using the following factors:

```
nullMean <- 0
popMean <- -1
popVar <- 1
sampleSize <- 4
alpha <- .05
tails <- 1
numReplications<- 100
```

a. What is the true probability of retaining the null hypothesis? __________
b. What is the true power? ___________
c. Why do you get these values for true probability and power? Hint: You are conducting a directional hypothesis, which only uses the probability area in one tail of the normal curve. ____
____

True or False Questions

Sampling Distributions

T F a. The sampling distribution of sample standard deviations is symmetrical for N = 2.
T F b. As sample size increases, the error in the sample standard deviation as an estimator of the population standard deviation decreases.
T F c. The sample standard deviation is an unbiased, efficient, consistent, and sufficient estimator of the population standard deviation.
T F d. The sample standard deviation tends to overestimate the population standard deviation.
T F e. On the average, the sample standard deviation is equal to the population standard deviation.
T F f. A consistent estimator is one that more closely approximates the population parameter as sample size increases.

Confidence Interval

T F a. A 95% confidence interval computed from sample data contains the population parameter approximately 95% of the time.
T F b. A confidence interval indicates the precision of a statistic to estimate the population parameter.
T F c. As sample size increases, the confidence interval width becomes narrower.
T F d. A 95% confidence interval implies that 95 times out of 100 the confidence interval *will not* contain the population parameter.
T F e. It is possible that a 95% confidence interval will not contain the population parameter.
T F f. As the confidence level increases from 90% to 99%, the width of the confidence interval becomes smaller.

Statistical Hypothesis

T F a. If the null hypothesis is false, it will be rejected approximately α percent of the time.
T F b. If the null hypothesis is false, it can still be retained.
T F c. The sample standard deviation may be used in place of the population standard deviation in the z-test if the sample size is small.
T F d. The probability of a TYPE II error increases if α is decreased.
T F e. If the p-value is greater than α, then the null hypothesis is rejected.
T F f. The region of rejection is denoted by a tabled statistical value for a given sample size and level of significance.
T F g. A directional hypothesis specifies the probability area for rejection of the null hypothesis in both tails of the normal distribution.
T F h. Different kinds of statistical hypotheses are tested depending on the level of measurement of data.
T F i. The level of significance (alpha) determines the probability area for the region of rejection.

TYPE I Error

T F a. The sum of the Confidence Level and TYPE I error probability equals 100%.
T F b. If the confidence level increases, the TYPE I error decreases.
T F c. As sample size increases, the confidence interval width decreases.
T F d. A TYPE I error is the probability of rejecting the Null Hypothesis when it is true and falsely accepting the Alternative Hypothesis.
T F e. The sample mean is always inside the confidence interval.
T F f. The population mean sometimes falls outside of the confidence interval.
T F g. With all other factors remaining constant, a 90% confidence interval would be narrower than a 99% confidence interval

TYPE II Error

T F a. As α increases, the power of the test increases.
T F b. Larger sample sizes will yield more powerful tests.
T F c. B is the probability of rejecting a true null hypothesis.
T F d. Small effect sizes are difficult to detect.
T F e. B decreases for larger sample sizes.
T F f. As sample size increases, power increases.
T F g. An acceptable value for power is .80.
T F h. Power is the probability of rejecting a null hypothesis that is false.
T F i. A TYPE I error refers to rejecting the null hypothesis when it is true.
T F j. A TYPE II error refers to retaining the null hypothesis when it is false.
T F k. Effect size refers to the difference between TYPE I and TYPE II errors.
T F l. A non-directional hypothesis indicates no difference between population parameters.
T F m. A directional hypothesis indicates that one population parameter is greater than the other population parameter.