Significance/Hypothesis Testing

A 6-Step Method for Tests of Significance

1. Model:
   a. Verbally identify the underlying random variable of interest.
   b. Verbally identify the underlying parameter of interest.
   c. State the assumptions being made about the underlying distribution.

2. State null and alternative hypotheses, H₀ and H₁.
   a. H₀, the null hypothesis, i.e., the hypothesis that is being tested, always includes a statement of equality. H₀ specifies a numerical "null value" of the parameter of interest, which will be presumed correct for the purpose of the test.
   b. H₁ (sometimes denoted by Hᴬ), the alternative, always includes a statement of strict inequality, and always contradicts H₀. H₁ determines whether the test is 2-sided or 1-sided. If possible, H₁ is a statement of the research hypothesis (Daniel 1987 p. 161, Daniel 1991 p. 191). This is
nearly always possible if the test is 1-sided. The conclusion, Step 6, is phrased in terms of $H_1$.

3. Formulate the test statistic.

4. Design the experiment or survey:
   a. Choose significance level, alpha
   b. Choose sample size, $n$.

5. Gather the data and analyze
   a. Compute the best point estimate of the parameter of interest.
   b. Compute the standard error of the estimate.
   c. Compute the observed value of the test statistic.
   d. Determine the $P$-value.
   e. Decide whether or not reject the null hypothesis in favor of the alternative.
      If $P < \alpha$, reject $H_0$. If $P > \alpha$, do not reject.
   f. Characterize the result verbally. If $P < \alpha$, "the results are significant." If $P > \alpha$, "not significant." If $P$ is an order of magnitude less than $\alpha$, the results are characterized as "highly significant".

6. State the conclusion verbally, succinctly, informatively:
   
   "There _____ significant statistical evidence that _______ ($\__ < P < \__$)."
   
   "There [is / is not] significant statistical evidence that [the alternative hypothesis, verbally] ($\__ < P < \__$)."

---

**Estimation versus Testing**

There are two kinds of statistical inference: estimation and testing. You are already familiar with estimation. You understand that the term "estimation" always means estimation of unknown parameters.

*Testing*, the second kind of inference, is also used to make inferences about the unknown value of a parameter. Scientists use testing rather than estimation when they are concerned about whether or not the unknown parameter has a specific value or a specific range of values.

To summarize and reiterate:

- In both kinds of inference, estimation and testing, there is an unknown parameter of interest, i.e., a parameter whose value is unknown.
- In testing, a specific value or range of values is of particular interest. That is, the scientist may have a theory that implies that the unknown parameter is
  - equal to some specific value,
  - not equal to some specific value,
  - less than some specific value, or
  - greater than some specific value.
The point is that the scientist uses testing when he or she has some testable, preconceived idea, i.e., hypothesis, about the unknown parameter of interest, and that hypothesis focuses interest on some specific value.

- When a scientist uses estimation he or she has no specific value in mind.

---

**Overview of Tests of Significance**

Significance testing, as described here, was introduced in the year 1900 by an English biomathematician named Karl Pearson. R. A. Fisher, another Englishman, championed and popularized this form of inference during the first quarter of this century.

**Purpose of Significance Testing**

R.A. Fisher (1956) put it this way: "As early as Darwin's experiments on growth rate the need was felt for some sort of a test whether an apparent effect might reasonably be due to chance."

**The Logic of Significance Testing**

Significance Testing is a form of statistical inference. A hypothesis, called the null hypothesis, is presumed to be true rather than some contradictory hypothesis, called the alternative hypothesis. Then relevant data are collected, and a test statistic is evaluated. The observed value of the test statistic is compared with what would be expected if the null hypothesis were true. If the observed value of the test statistic is "far" from what would be expected if the null hypothesis were true, then the data contradict the null hypothesis and support the alternative hypothesis and the result is called (statistically) significant. If, on the other hand, the observed value of the test statistic is not "far" from what would be expected if the null hypothesis were true, then the data do not contradict the null hypothesis, do not support the alternative hypothesis, and the result is called not (statistically) significant. If the data are “close” to (i.e., not “far” from) what would be expected under the null hypothesis, the results are considered inconclusive. That is, we cannot infer that the null hypothesis is true just because the data agree with the null hypothesis. The data might be explained just as well by some other hypothesis. On the other hand, if the data are "far" from agreeing with the null hypothesis, then we do infer that the null hypothesis is implausible, and that the alternative must therefore probably be true. This is a simple, but very important, matter of logic. Moreover, it is the reason that we call results that refute the null hypothesis statistically significant, and results that fail to refute the null hypothesis not statistically significant.

The logic of significance testing is the logic of Galileo's scientific method. Karl Pearson's contribution is the device used to quantify the distance between the data and what would be expected if the null hypothesis were true. That device is called the $P$-value.
Fisher (1956) put it this way: “The logical basis of these scientific applications was the elementary one of excluding, at an assigned level of significance [i.e., $P$-value], [null] hypotheses, or views of causal background, which could only by a more or less implausible coincidence have led to what had been observed.” $P$-values will be explained after we look at an example.

**Tests of Hypothesis**

Hypothesis testing was invented in 1933 by two of Karl Pearson's students, Jerzy Neyman and (Karl’s son) Egon Pearson. Most statisticians refer to hypothesis testing as the Neyman-Pearson Method. Many statisticians, unlike me, do not think that the difference between significance testing and hypothesis testing is very important; consequently, the two methods are often confused.

In hypothesis testing, the underlying model (Step 1), the null and alternative hypotheses (Step 2), and the test statistic (Step 3), are formulated exactly the same way as in significance testing.

The first difference arises in the design step (Step 4), when it becomes necessary to choose a *significance level*, $\alpha$, which will be defined soon.

*After* choosing the level of significance, the investigator chooses the sample size. Then (Step 5) relevant data are collected, and, based on these data, a *decision* is made. It is decided either

- to REJECT the null hypothesis in favor of the alternative, or
- NOT to REJECT the null hypothesis in favor of the alternative.

The following two types of error can result:

Type 1 error:
  - to REJECT the null hypothesis when the null hypothesis actually is true, or

Type 2 error:
  - NOT to REJECT the null hypothesis when the null hypothesis actually is false.
Table 1  Truth Table. The Truth Table defines Type 1 Error, Type 1 Error Rate ($\alpha$), Type 2 Error, Type 2 Error Rate ($\beta$), and Power ($1 - \beta$).

<table>
<thead>
<tr>
<th>Truth Table</th>
<th>(Unknown) True State of Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$, Results significant, There is evidence in favor of $H_1$.</td>
<td>$H_0$ is true</td>
</tr>
<tr>
<td>Correct decision, $\alpha = \text{Level of Significance} = \text{Type 1 Error Rate}$</td>
<td>Type 1 Error, $(1 - \beta) = \text{Power}$</td>
</tr>
<tr>
<td>NOT Reject $H_0$, Results NOT Significant, There is NOT evidence in favor of $H_1$.</td>
<td>Correct decision $(1 - \alpha)$</td>
</tr>
</tbody>
</table>

The first priority is to avoid a Type 1 error, and the null and alternative hypotheses are chosen with this in mind.

The significance level of the test is denoted by $\alpha$, and is defined as the probability of committing a Type 1 error, i.e., the probability of rejecting the null hypothesis given that the null hypothesis is true. It is by setting the significance level, $\alpha$, at a small value like 0.001 or 0.01 or 0.05 or 0.10 that we control the chance of erroneously REJECTing a true null hypothesis (in favor of a false alternative). Thus the investigator controls the chance of committing a Type 1 error, which is the first priority.

Note. The result of a significance test is a $P$-value which is a statistic that quantifies the plausibility of the null hypothesis with respect to the alternative. The result of a hypothesis test is a decision, either to reject, or not to reject, the null hypothesis in favor of the alternative.

The two methods are easily reconciled by thinking of a hypothesis test as a significance test with a few augmentations. First, before gathering the data, the hypothesis tester chooses a significance level $\alpha$ which will be the largest $P$-value he or she will take as support of the alternative hypothesis. After gathering the data, the hypothesis tester computes the $P$-value, and compares the observed $P$-value with the predetermined $\alpha$-level. If the observed $P$-value is smaller than the predetermined $\alpha$-level, the hypothesis tester rejects the null hypothesis in favor of the alternative. If, on the other hand, the observed $P$-value is larger than the predetermined $\alpha$-level, then the hypothesis tester does not reject the null hypothesis in favor of the alternative.

**The Jury-Trial/Hypothesis-Testing Analogy**

The logic of hypothesis testing is analogous to the logic of our system of jurisprudence, where the defendant is presumed innocent until proven guilty beyond a shadow of a doubt.
### Truth Table

<table>
<thead>
<tr>
<th>Jury Trial</th>
<th>(Unknown) True State of Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H_0$ is true</td>
</tr>
<tr>
<td></td>
<td><em>Innocent</em></td>
</tr>
<tr>
<td>Decision</td>
<td>Reject $H_0$, Results significant, There is evidence in favor of $H_1$. <em>We find the defendant guilty.</em></td>
</tr>
<tr>
<td></td>
<td>NOT Reject $H_0$, Results NOT Significant, There is NOT evidence in favor of $H_1$. <em>We find the defendant NOT guilty</em></td>
</tr>
</tbody>
</table>

In a jury trial, the presumption of innocence is the "null hypothesis". The "alternative hypothesis" is that the accused is guilty. The "data" are the bits of evidence presented to the jury during the trial, and the decision made by the jury based on the evidence (data) is either

- we find the defendant GUILTY (i.e., we REJECT the null hypothesis in favor of the alternative), or
- we find the defendant NOT GUILTY (i.e., we do NOT REJECT the null hypothesis in favor of the alternative).

The jury can make the following two types of error:

**Type 1 error:**
find the defendant guilty when he actually is innocent, or

**Type 2 error:**
find the defendant not guilty when he actually is guilty.

As you know, the *first priority* is to avoid a "Type 1 error," convicting an innocent man. In a jury trial this is controlled by choosing a very small "significance level" and a shadow of a doubt.

In a jury trial, if the jury becomes convinced that the defendant is innocent, then they find him NOT GUILTY, i.e., they do NOT REJECT the "null hypothesis" that he is innocent in favor of the "alternative" that he is guilty. But--and this is very important--it is not necessary for the jury to be convinced beyond a shadow of a doubt that the defendant is innocent in order to find that he is NOT GUILTY. As long as they fail to be convinced beyond a shadow of doubt that the defendant is guilty, they should find him NOT GUILTY. On the other hand, in order to find the defendant GUILTY, the jury must be convinced beyond a shadow of a doubt that he is guilty. Thus, a finding of NOT GUILTY
does *not* imply that the jury is convinced that the defendant is innocent, it implies only that they failed to be convinced, beyond a shadow of a doubt, that he is guilty. On the other hand, a finding of GUILTY does imply that the jury is convinced beyond a shadow of a doubt that the defendant is GUILTY.

Analogously, in hypothesis testing, if the investigator decides NOT to REJECT the null hypothesis in favor of the alternative, that does *not* imply that he is convinced that the null hypothesis is true. It implies only that the researcher failed to be convinced that the null hypothesis is false and that he failed to be convinced that the alternative is true. On the other hand, if the researcher does decide to REJECT the null hypothesis in favor of the alternative, then it is implied that he is convinced that the null hypothesis is false and that the alternative is true.

---

**Example A: Significance Test, using the 6-step method**

(6.2.1, p. 194 of Daniel 1987) Researchers are interested in whether the mean level of some enzyme is abnormal among adult white male alcoholics. The normal level is 25, with a population standard deviation of 6.71. Based on a sample of 10 adult male alcoholics with a mean enzyme level of 22, is there evidence that the mean enzyme level for adult male alcoholics is abnormal?

**The 6-Step Method of Significance Testing**

1. **Model**
   a. Let $X_i$ be the enzyme level for the $i^{th}$ randomly sampled adult male alcoholic (ama).
   b. Let $E(X_i) = \mu = $ Population Mean = the mean level for all amas.
   c. Assume
      i. Population standard deviation $\sigma = \sqrt{45} = 6.71$
      ii. $X_i$ distributed normally.

2. **Hypotheses**
   $H_0$: $\mu = 25$, versus $H_1$: $\mu \neq 25$

3. **Test Statistic**

   $$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

   (1)
where $\bar{X}$ denotes the sample mean, and

where $\mu_0$ denotes the hypothetical value of the population mean.

[We use formula (1) because the parameter of interest is the population mean, and the population variance is (assumed to be) known.]

4. Design
   a. $\alpha = 0.05$
   b. $n = 10$

5. Gather Data
   a. Best estimate of population mean = sample mean = $\bar{X} = 22$
   b. $SE = \sigma_{\bar{X}} = \sigma/\sqrt{n} = 6.71/\sqrt{10} = 2.12$
   c. Test Statistic
      $Z_{OBS} = (22 - 25)/2.12 = -1.41$
   d. P-value: Since $H_1$ specifies $\neq$,
      $P = 2P\{Z \geq |Z_{OBS}|\} = 2P\{Z \geq |-1.41|\} = 0.1586$

      (from a detailed Normal table)

      Alternatively, we can use a t-table by realizing that the bottom row of the t-table, corresponding to infinite degrees of freedom, is a row of standard normal critical values. We look for the critical values adjacent to $|Z_{OBS}| = 1.41$ and find that $0.10 < P < 0.20$. 

<table>
<thead>
<tr>
<th>Standard Normal</th>
<th>= Infnite</th>
<th>1.282</th>
<th>1.645</th>
<th>1.960</th>
<th>2.326</th>
<th>2.576</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-value for one-sided alternative:</td>
<td>0.10</td>
<td>0.05</td>
<td>0.025</td>
<td>0.01</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>P-value for two-sided alternative:</td>
<td>0.20</td>
<td>0.10</td>
<td>0.05</td>
<td>0.02</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Confidence level (central area):</td>
<td>0.80</td>
<td>0.90</td>
<td>0.95</td>
<td>0.98</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Percentile rank (probability of a lesser value):</td>
<td>0.90</td>
<td>0.95</td>
<td>0.975</td>
<td>0.99</td>
<td>0.995</td>
<td></td>
</tr>
</tbody>
</table>

e. Decision: Not reject $H_0$

f. Characterize the results as follows.
   - The results are not significant ($P > 0.10$), or
• The results are not significant ($0.10 < P < 0.20$), or
• The results are not significant ($P = 0.16$).

6. Conclusion
   There is not statistically significant evidence that the mean enzyme level of adult male alcoholics is different from 25 ($P > 0.10$).

Scientific Method

The beauty of significance testing is the simplicity of its logic. The logic is not due to Karl Pearson; it is at least as old as the scientific method of Galileo (1564-1642). (Newton was born the year Galileo died.) The steps of the scientific method are

1. Formulate a hypothesis.
2. Deduce some consequences of the theory that should be observable.
3. Test the hypothesis by performing an experiment or by otherwise making the observations.
   a. If the observed results are inconsistent with hypothesis, then the hypothesis has been refuted, and scientific progress has been made. Go back to step one and formulate a new hypothesis.
   b. If the results of the experiment are consistent with the hypothesis, then the hypothesis has not been refuted, it remains viable, but it hasn't been proven. The hypothesis can never be proven because
      ▪ The consequence of the hypothesis that was tested might also be the consequence of some other hypothesis.
      ▪ There might be some other consequence of the hypothesis that will not pass the test of observation.
      ▪ Above all, the hypothesis can not be proven as a matter of logic.

The investigator has several choices now.

1. Devise a new and different test of the consequence of Step 2, and go back to Step 3.
2. If the consequence of Step 2 has been tested in many ways, derive a new and different consequence and go back to Step 2.
3. If many consequences of the hypothesis have been derived, if they've all been tested in many ways and passed, and if you can't think up any new and different consequence or tests, then proclaim the hypothesis a theory, formulate a new and different hypothesis, and go back to Step 1.
Details of the Six Step Method

Step 1: The Model

The modeling step is the same as in confidence interval estimation.

Step 2: Specifying the Hypotheses

The statistical hypotheses are always statements about the value of the parameter of interest.

The Null Hypothesis

- $H_0$, the null hypothesis, i.e., the hypothesis that is being tested, always includes a statement of equality.
- $H_0$ specifies a numerical "(null value)" of the parameter of interest, which will be presumed correct for the purpose of the test.

The Alternative Hypothesis

- Some authors denote the alternative hypothesis by $H_1$, others by $H_A$, or $H_a$.
- $H_1$, the alternative, always includes a statement of strict inequalities: either $<$, $>$, or $\neq$.
- $H_1$ always contradicts $H_0$.
- $H_1$ partially determines how to compute the $P$-value.
- If possible, $H_1$ is a statement of the research hypothesis.
- The conclusion, Step 6, is always phrased in terms of $H_1$.

The key to proper specification is close adherence to these "rules", especially the rule that the conclusion is always a statement about the alternative, $H_1$.

The conclusion always takes the form:

There [is / is not] statistically significant evidence that ____________ $(? < P < ?)$.

The blank is always filled in by a verbal statement of the alternative hypothesis.

Note that the null and alternative hypotheses are a pair. If we denote the parameter of interest by $\theta$ and the null value by $\theta_0$, then the three possible pairs are

- $H_0$: $\theta = \theta_0$ vs. $H_1$: $\theta < \theta_0$, called a one-sided lower-tailed test,
- $H_0$: $\theta = \theta_0$ vs. $H_1$: $\theta > \theta_0$, called a one-sided upper-tailed test,
- $H_0$: $\theta = \theta_0$ vs. $H_1$: $\theta \neq \theta_0$, called a two-sided or two-tailed test.
Some authors prefer to denote these three pairs of hypotheses by

$$H_0: \theta \geq \theta_0 \text{ vs. } H_1: \theta < \theta_0,$$
called a one-sided lower-tailed test,

$$H_0: \theta \leq \theta_0 \text{ vs. } H_1: \theta > \theta_0,$$
called a one-sided upper-tailed test,

$$H_0: \theta = \theta_0 \text{ vs. } H_1: \theta \neq \theta_0,$$
called a two-sided test or two-tailed test.

I prefer the former because it emphasizes that for the purpose of the test, the sample estimator of $\theta$ is going to be compared with the specific value of the population parameter $\theta = \theta_0$. Either notational convention is valid, as they mean the same thing.

**Summary of Step 2, Stating the Null and Alternative Hypotheses**

<table>
<thead>
<tr>
<th>Research (Biological) Hypothesis</th>
<th>Null Hypothesis</th>
<th>Alternative Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta &lt; \theta_0$</td>
<td>$H_0: \theta = \theta_0$</td>
<td>$H_1: \theta &lt; \theta_0$</td>
</tr>
<tr>
<td>$\theta &gt; \theta_0$</td>
<td>$H_0: \theta = \theta_0$</td>
<td>$H_1: \theta &gt; \theta_0$</td>
</tr>
<tr>
<td>$\theta \neq \theta_0$</td>
<td>$H_0: \theta = \theta_0$</td>
<td>$H_1: \theta \neq \theta_0$</td>
</tr>
<tr>
<td>$\theta = \theta_0$</td>
<td>$H_0: \theta = \theta_0$</td>
<td>$H_1: \theta \neq \theta_0$</td>
</tr>
</tbody>
</table>

**Conclusion (Step 6)**

The alternative hypothesis partially determines the conclusion of Step 6.

“*There [is/is not] significant statistical evidence that [H1] (P = [__]).*”

- The alternative does not determine whether there is or is not significant evidence. We need the data (step 5) to determine that.
- The alternative does not determine whether the P-value. We need the data (step 5) to determine that.
- The alternative does determine what the conclusion will address. The conclusion addresses whether or not there is significant statistical evidence in favor of the alternative. The conclusion is always about the alternative; never about the null.

**Type 1 and Type 2 Error (Interpretation of the Conclusion, Beyond Step 6)**

Specifying the hypotheses automatically determines what constitutes a Type 1 and Type 2 error for the particular research hypothesis. Recall that Type 1 and Type 2 error are defined in the Truth Table, Table 1 on page 5.
Exercises on stating hypotheses

For each of the following research objectives, perform the following steps of the 6-Step Method, and state what constitutes a Type 1 and Type 2 Error.

[Step 1a] Verbally identify the random variable.

[Step 1b] Verbally identify the parameter of interest (as in the modeling step).

[Step 2] Formulate the null and alternative hypothesis symbolically.

[Step 6b] Check your formulation by

(i) State the Conclusion/Results.
(ii) State what constitutes a Type 1 error in terms of the parameter of interest and the hypotheses.
(iii) State what constitutes a Type 2 error in terms of the parameter of interest and the hypotheses.

Exercise (1a) is done as an example.

1. First consider the questions we could ask about the mean enzyme level (mg/dL) of adult male alcoholics.
   a. Is there evidence that the mean enzyme level (mg/dL) is not 25?
   b. Is there evidence that the mean enzyme level (mg/dL) is less than 25?
   c. Is the mean enzyme level greater than 25 mg/dL?
   d. Is the mean enzyme level 25 mg/dL?

2. As part of our job in the Montgomery County Division of Social Services, we are considering the idea of petitioning the Department of Health and Human Services for funds made available to qualifying communities by a recent act of Congress. The Department of Health and Human Resources must grant the funds if the mean family income in Montgomery County is less than $20,000. Do we have a case?

3. We are entitled to a grant if more than 20% of the families have an income of less than $20,000. Are we entitled to a grant? (Hint: The random variable of interest is categorical.)

4. According to Gregor Mendel's theory, 50% of the F1 generation should have white eyes. Test Mendel's theory.

5. Einstein claims that the speed of light is 30 billion cm/sec. Test Einstein's claim.
Step 3: Formulate the Test Statistic

A statistic is a random variable that is a function of the sample observations and that depends on no unknown parameters. The test statistic reduces the entire sample to a single number that determines the outcome of the test. As a result of many years of research, statisticians have compiled a long list of test statistics, each designed for a different type of experiment or survey. Initially, we will learn how to apply three of these test statistics.

The choice of the proper test statistic is based on the model, particularly

- the parameter of interest and
- the assumptions made about the underlying distribution.

1. If
   a. the parameter of interest is a population mean ($\mu$), and
   b. the underlying population standard deviation ($\sigma$) is (assumed to be) known, and
   c. either
      i. the underlying random variable of interest is (assumed to be) normally distributed, or
      ii. the sample size is large (enough that the CLT applies),

   then the appropriate test statistic is

   $$ Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} , $$

   where $\bar{X}$ denotes the sample mean and $\mu_0$ denotes the null value that will be assigned to the parameter of interest $\mu$ in the null hypothesis $H_0 : \mu = \mu_0$

2. If
   a. the parameter of interest is a population mean, $\mu$, and
   b. the underlying population standard deviation ($\sigma$) is unknown, i.e., (not assumed to be known), and
   c. the underlying random variable of interest is (assumed to be) normally distributed,

   then the appropriate test statistic is Student’s (1908) $T$ statistic defined by

   $$ T_{n-1} = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} , $$

   where $\bar{X}$ denotes the sample mean and $\mu_0$ denotes the null value that will be assigned to the parameter of interest $\mu$ in the null hypothesis $H_0 : \mu = \mu_0$
Note. Use of $T$ always requires the assumption that the underlying distribution is normal.

3. If
   a. the parameter of interest is a population standard deviation, and
   b. the underlying random variable of interest is (assumed to be) normally distributed,

then the appropriate test statistic is Karl Pearson’s (1900) *chi-squared statistic* with $n - 1$ degrees of freedom, denoted by $\chi^2_{n-1}$ and defined by

$$\chi^2_{n-1} = (n - 1) \frac{s^2}{\sigma_0^2}$$

where $n$ denotes the sample size, $s$ denotes the sample standard deviation, $s^2$ denotes the sample variance, and $\sigma_0^2$ denotes the null hypothetical value of the population variance.

**Assumption.** Use of the chi-squared statistic, $\chi^2_{n-1}$, for inference about the population standard deviation, $\sigma$, always requires the assumption that the underlying distribution is normal.

4. If the parameter of interest is a population proportion, $\varphi$, then the appropriate test statistic is

$$Z = \frac{p - \varphi_0}{\sqrt{\varphi_0(1-\varphi_0)/n}}$$

where $p$ denotes the sample proportion, $\varphi_0$ denotes the null hypothetical value of the parameter of interest, the population proportion.

**Step 4: Design**

Choose the significance level, $\alpha$ (*alpha*)

After the study is performed and the data are analyze, a statistic called the *P-value* will be calculated under the assumption that the null hypothesis is true. The P-value can be thought of as the probability of the sample given that the null hypothesis is true. The result of the significance/hypothesis test will be determined by the magnitude of the P-value. The significance level, $\alpha$, draws the line between a small P-value and a large P-value. The P-value is denoted by $P$. 

• If $P \leq \alpha$, then the probability of the sample is very small if the null hypothesis is true, therefore
  o the decision is to reject the null hypothesis in favor of the alternative (i.e., accept the alternative)
  o the results are characterized as statistically significant,
  o the conclusion is, “There is significant statistical evidence in favor of the alternative hypothesis.”

• If $P > \alpha$, then the probability of the sample is not very small if the null hypothesis is true, therefore
  o the decision is to not reject the null hypothesis in favor of the alternative (i.e., not accept the alternative)
  o the results are characterized as not statistically significant,
  o the conclusion is, “There is not significant statistical evidence in favor of the alternative hypothesis.”

• If the null hypothesis is true, then the decision to not reject is correct, while the decision to reject is erroneous: a Type 1 Error. The significance level $\alpha = \text{the probability of rejecting } H_0 \text{ (erroneously), given that } H_0 \text{ is true.}$

• Note that $\alpha$, the Type 1 Error Rate, is not the unconditional probability of committing a Type 1 Error. A Type 1 Error can be committed only if $H_0$ is true. Thus, the significance level $\alpha = \text{the conditional probability of rejecting } H_0 \text{ (erroneously), given that } H_0 \text{ is true.}$

• More about Type 1 and Type 2 Error later.

Choose the sample size, $n$. The sample size partially determines the Type 2 Error Rate and the Power of the test.

**Sensitivity, i.e., Power**

The larger the sample size, the more information (data) we have about the underlying population or experimental phenomena. More information means a more sensitive test, i.e., more sensitive to small departures (of the true, actual value of the parameter of interest) from the null hypothetical value. A more sensitive test has a greater probability of rejecting the null hypothesis in favor of the alternative when the true value of the parameter of interest differs only slightly from the null hypothetical value. An insensitive test has a high probability of rejecting (the null in favor of the alternative) only when there is a large difference between the (null value) and the true value.
Step 5: Gather the data and Analyze

As in confidence interval estimation, we compute the best point estimate of the parameter of interest and (except when making inferences about the population SD) the standard error of the estimate. With these and the (null value) of the parameter of interest, we can then compute the observed value of the test criterion (statistic). The observed value of the test statistic, be it $Z_{OBS}$ or $T_{OBS}$ is then used to determine the P-value as follows.

Step 5d: P-value

Definition of the P-value

For purposes of illustration, suppose that the parameter of interest is a population mean (a population proportion would do just as well), and that the null hypothesis is $H_0$: (population mean) = (null value), and suppose that the test statistic is $Z$, and denote the observed value of $Z$ by $Z_{OBS}$.

- If the alternative hypothesis is $H_1$: (population mean) < (null value), then the P-value is defined by

  \[ P = P\{Z < Z_{OBS}\} \]

- If the alternative hypothesis is $H_1$: (population mean) > (null value), then the P-value is defined by

  \[ P = P\{Z > Z_{OBS}\}. \]
- If the alternative hypothesis is \( H_1: \) (population mean) \( \neq \) (null value), then the P-value is defined by

\[
P = 2 \times P(Z > |Z_{OBS}|).
\]

Note that all of these probabilities are computed under the presumption that the null hypothesis is true. Thus, the P-value is defined as the probability, presuming that the null hypothesis is true, of observing a value of the test statistic (\( Z \), for example) as "extreme" or more extreme than its observed value (\( Z_{OBS} \), for example), in the direction specified by the alternative hypothesis.
## Determination of the P-value

- If the test statistic is \( Z \) (standard normal), and if a detailed table of the Standard Normal CDF is available, then the P-value is determined by means of the definition above.

- If the test statistic is \( Z \) (standard normal) but only a \( t \)-table is available, then the P-value is determined as follows.
  
  **Example 1:** Suppose you are testing \( H_0: \) population mean = 150 vs. \( H_1: \) (population mean) > 150, and suppose you have observed \( Z = 2.2 \). First, this is a one-sided upper-tailed test, therefore \( P = P\{Z > 2.2\} \).
  
  Second, because the observed value of 2.2 is greater than the expected value of 0 (recall that the expected value of a \( Z \) statistic is 0), the observed value of the test statistic is in the upper tail. It’s a good idea to sketch the PDF and shade in the P-value:

![PDF Sketch](image)

From JMP or Minitab: \( P = P\{Z > 2.2\} = 0.0139 \approx 1.4\% \)

From the bottom line of the \( t \)-table for infinite degrees of freedom we find

<table>
<thead>
<tr>
<th>( P )-value for one-sided alternative:</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P-value for two-sided alternative:</strong></td>
<td>0.20</td>
<td>0.10</td>
<td>0.05</td>
<td>0.02</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Confidence level (central area):</th>
<th>0.80</th>
<th>0.90</th>
<th>0.95</th>
<th>0.98</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Percentile rank</strong> (probability of a lesser value):</td>
<td>0.90</td>
<td>0.95</td>
<td>0.975</td>
<td>0.99</td>
<td>0.995</td>
</tr>
</tbody>
</table>

Because \( Z_{OBS} = 2.2 \) is between 1.96 and 2.326, the one-sided upper-tailed P-value is

\( 0.01 < P < 0.025 \)
Example 2: Suppose you are testing $H_0$: (population mean) $= 150$ vs. $H_1$: (population mean) $< 150$, and suppose you have observed $\{Z = -1.7\}$. Since this is a lower-tailed test, $P = P\{Z < -1.7\}$. From the bottom line of the t-table for infinite degrees of freedom we find $-1.96 < -1.7 < -1.645$, therefore $0.025 < P < 0.05$. First, this is a one-sided lower-tailed test, therefore $P = P\{Z < -1.7\}$. Second, because the observed value of $-1.7$ is less than the expected value of 0 (recall that the expected value of a Z statistic is 0), the observed value of the tests statistic is in the lower tail.

![Standard Normal Distribution Image]

From JMP or Minitab: $P = P\{Z < -1.7\} = 0.0446 \approx 4.5\%$

From the bottom line of the t-table for infinite degrees of freedom we find

<table>
<thead>
<tr>
<th>P-value for one-sided alternative:</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-value for two-sided alternative:</td>
<td>0.20</td>
<td>0.10</td>
<td>0.05</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Confidence level (central area):</td>
<td>0.80</td>
<td>0.90</td>
<td>0.95</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>Percentile rank (probability of a lesser value):</td>
<td>0.90</td>
<td>0.95</td>
<td>0.975</td>
<td>0.99</td>
<td>0.995</td>
</tr>
</tbody>
</table>

Because $Z_{OBS} = -1.7$ is between $-1.96$ and $-1.645$, the one-sided lower-tailed P-value is

$0.025 < P < 0.050$
Example 3: Suppose you are testing $H_0$: (population) mean $= 150$ vs. $H_1$: (population mean) $\neq 150$, and suppose you have observed $\{Z = -1.7\}$. This is a two-tailed test, therefore

$$P = 2 \, P\{Z > |1.7|\}$$

which is the fraction of the following Standard Normal PDF that is shaded in the following graph.

From JMP, Minitab, etc., we get $P = 2 \times 0.0446 = 0.0892$.

<table>
<thead>
<tr>
<th>Standard Normal</th>
<th>Infinite</th>
<th>1.282</th>
<th>1.645</th>
<th>1.960</th>
<th>2.326</th>
<th>2.576</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-value for one-sided alternative:</td>
<td></td>
<td>0.10</td>
<td>0.05</td>
<td>0.025</td>
<td>0.01</td>
<td>0.005</td>
</tr>
<tr>
<td>P-value for two-sided alternative:</td>
<td></td>
<td>0.20</td>
<td>0.10</td>
<td>0.05</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Confidence level (central area):</td>
<td></td>
<td>0.80</td>
<td>0.90</td>
<td>0.95</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>Percentile rank (probability of a lesser value):</td>
<td></td>
<td>0.90</td>
<td>0.95</td>
<td>0.975</td>
<td>0.99</td>
<td>0.995</td>
</tr>
</tbody>
</table>

From the bottom line of the t-table for infinite degrees of freedom we see

$$1.645 < 1.7 < 1.96,$$

therefore

$$0.05 < P < 0.10.$$
Example 4: Suppose you are testing

\[ H_0: \mu = 150 \text{ vs. } H_1: \mu > 150, \text{ where } \mu \text{ denotes the population mean.} \]

Suppose you have observed

\( \{Z = -1.7\} \).

Because this is an upper-tailed test,

\[ P = P\{Z > -1.7\}. \]

Without looking in any table, we know that

\[ P\{Z > 0\} = 0.5, \]

therefore

\[ P\{Z > -1.7\} > 0.5, \]

and no one would consider the results significant. In a situation like this, it would suffice to report:

“The results were not significant (P > 0.10).”

Another way to see this is to realize that if \( Z_{OBS} \) is negative, i.e., in the lower tail, then the estimator, the observed sample mean, was less than the null hypothetical value of 150. If the sample mean is less than 150, then we wouldn’t make sense to conclude that the population mean is greater than 150. So we know the result cannot be statistically significant. See the “Interpretation of P-values” below.
• If the test statistic is **Student's T** with \((n - 1)\) degrees, \(T_{n-1}\) of freedom, then use the t-table in the same way, but use \((n - 1)\) instead of infinite degrees of freedom.

• If the test statistic is **Chi-squared** with \((n - 1)\) degrees of freedom,

\[
\chi^2_{n-1} = (n-1) \frac{s^2}{\sigma_0^2}
\]

then the P-value is determined as follows. (Let \(\sigma\) denote the population standard deviation, and let \(s\) denote the sample standard deviation.)

- if the alternative hypothesis is \(H_i: \sigma < \sigma_0\), then \(P = P\{\chi^2_{n-1} \leq \chi_{OBS}^2\}\).
- if the alternative hypothesis is \(H_i: \sigma > \sigma_0\), then \(P = P\{\chi^2_{n-1} \geq \chi_{OBS}^2\}\).
- if the alternative hypothesis is \(H_i: \sigma \neq \sigma_0\), then the determination of the P-value depends on which tail the observed sample standard deviation \(s\) lies in.
  - if the alternative is \(H_i: \sigma \neq \sigma_0\), and if \(s/\sigma_0 < 1\), then \(P = 2P\{\chi^2_{n-1} \leq \chi_{OBS}^2\}\).
  - if the alternative is \(H_i: \sigma \neq \sigma_0\), and if \(s/\sigma_0 > 1\), then \(P = 2P\{\chi^2_{n-1} \geq \chi_{OBS}^2\}\).

- **Example 5**: Suppose you are testing

\(H_0: \sigma = 50\) vs. \(H_1: \sigma > 50\), where \(\sigma\) denotes the population standard deviation.

Suppose you have observed a sample standard deviation of \(s = 62\) with 15 degrees of freedom based on a sample of \(n = 16\) observations.

Then the observed value of the test criterion is

\[
\chi^2_{15} = (n-1) \frac{s^2}{\sigma_0^2} = (15) \frac{62^2}{50^2} = 23.1
\]

Because this is an upper-tailed test,

\[
P = P\{\chi^2_{15} > \chi_{OBS}^2\} = P\{\chi_{15}^2 > 23.1\} = 0.082
\] (1)
The calculation of Equation (1) and the diagram were computed by means of JMP script ChiSqPDFandCDF.JSL.

Chi-Squared P-values can also be computed within a JMP data table as P-Values.JMP.

To calculate the P-value from a tabulation of Chi-Squared critical values, the following thought process is required.

We must first notice that the observed value of the test statistic, \( \chi^2_{15} = 23.1 \), is in the upper tail because \( \chi^2_{15} = 23.1 > 15 = E\left(\chi^2_{15}\right) \). Consulting the table we find \( \chi^2_{15,0.90} = 22.307 \) and \( \chi^2_{15,0.95} = 24.996 \). Thus, the observed \( \chi^2_{15} = 23.1 \) is somewhere between the 90th and 95th percentile, therefore the upper-tail probability is between 0.10 and 0.05, so

\[
0.05 < P < 0.10
\]

Because \( a = 0.05 \) and \( P > 0.05 \), we have \( P > a \), so the decision is to

“NOT reject the null hypothesis
at the 0.05 level of significance”

We characterize the results by

“The results were NOT significant (0.05 < P < 0.10)”

or
“The results were NOT significant (P = 0.08)”

And we conclude that

“There is NOT significant statistical evidence that the population standard deviation is greater than 50 (P > 0.10).

or

“There is NOT significant statistical evidence that the population standard deviation is greater than 50 (P = 0.08).

If the alternative had been

\[ H_1: \sigma < 50 \]

and the observed value of the test criterion were

\[ \chi^2_{15} = 23.1 \]

then the P-value would be

\[ P = P\{ \chi^2_{n-1} < \chi^2_{\text{obs}} \} = P\{ \chi^2_{15} < 23.1 \} = 0.918 \]

The decision would be NOT to reject the null hypothesis, the results would be characterized as NOT significant, and the conclusion would be that there is NOT significant statistical evidence that the population standard deviation is less than 50.

for the two-sided alternative

\[ H_1: \sigma \neq 50 \]

and the observed test criterion of

\[ \chi^2_{15} = 23.1 \]

then the P-value would be

\[ P = 2P\{ \chi^2_{15} > 23.1 \} = 2(0.082) = 0.164 \] (2)

The P-value is 2 times the upper-tail probability in this case because the observed \( \chi^2_{15} = 23.1 \) happened to be in the upper tail. If the observed value of
had been less than the degrees of freedom of 15 and therefore in the lower tail, then the P-value would be 2 times the lower-tail probability.

See the “Interpretation of P-values” below.

**Exercise on Determination of P-Value**

**Instructions.** You can use a JMP Probability function in the formula editor to calculate P exactly, and then report the answer in the form \( P = a \). Alternatively, you can use a statistical table to find a range for the P-value, and report it in the form \( b < P < c \).

1. In testing \( H_0: \mu = 25 \) vs. \( H_1: \mu \neq 25 \), an investigator observed \( T = 1.6 \) with 12 degrees of freedom.
   a. What is the P-value?
   b. What would the P-value be if the alternative had been \( H_1: \mu > 25 \)?
   c. What would the P-value be if the alternative had been \( H_1: \mu < 25 \)?
2. In testing \( H_0: \varphi = 0.75 \) vs. \( H_1: \varphi \neq 0.75 \), an investigator observed \( Z = 3.2 \).
   a. What is the P-value?
   b. What would the P-value be if the alternative had been \( H_1: \varphi > 0.75 \)?
   c. What would the P-value be if the alternative had been \( H_1: \varphi < 0.75 \)?
3. In testing \( H_0: \sigma = 25 \) vs. \( H_1: \sigma \neq 25 \), an investigator observed \( \chi^2_{n-1} = 19.1 \) with 12 degrees of freedom.
   a. What is the P-value?
   b. What would the P-value be if the alternative had been \( H_1: \sigma > 25 \)?
   c. What would the P-value be if the alternative had been \( H_1: \sigma < 25 \)?

**Interpretation of P-values**

There are several mutually consistent ways to interpret a P-value:

- The **smaller** the P-value, the greater is the “strength of the evidence (the data, the sample)” against the null hypothesis and in favor of the alternative.

- The **smaller** the P-value, the greater is the “distance between the null hypothesis and the data.”

- R. A. Fisher (1956, p. 39) put it this way: “The force with which such a conclusion [i.e., a conclusion, based on a small P-value, that ‘the result is significant’] is supported is logically that of the simple disjunction: Either an exceptionally rare chance has occurred, or the theory [which implies the null hypothesis] is not true.”
Step 5e Decision, Verbal characterization of the Results, and
Step 6, Verbal Conclusion

There are three ways of characterizing the “statistical significance” of the results based on the P-value.

In significance testing, as devised by Karl Pearson in 1900, and as popularized by R.A. Fisher with his 1925 test book, and before the introduction of the $\alpha$-level (significance level) by Neyman and Pearson (1933) the result was traditionally characterized verbally as follows.

<table>
<thead>
<tr>
<th>P-value: Verbal characterization of the observed P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P &lt; 0.001$: The results are very highly significant ($P &lt; 0.001$).</td>
</tr>
<tr>
<td>$0.001 &lt; P &lt; 0.01$: The results are highly significant ($0.001 &lt; P &lt; 0.01$).</td>
</tr>
<tr>
<td>$0.01 &lt; P &lt; 0.05$: The results are significant ($0.01 &lt; P &lt; 0.05$).</td>
</tr>
<tr>
<td>$0.05 &lt; P &lt; 0.10$: The results are slightly significant ($0.05 &lt; P &lt; 0.10$).</td>
</tr>
<tr>
<td>$P &gt; 0.10$: The results are NOT significant ($P &gt; 0.10$).</td>
</tr>
</tbody>
</table>

The verbal characterization of significance is subjective, but since the P-value is given, the reader is free to make his or her own subjective assessment.

In hypothesis testing, as devised by Jersey Neyman and Egon Pearson in 1928-33, a decision, whether or not to reject the null hypothesis in favor of the alternative, is made based on comparison of the P-value with significance level $\alpha$ that was specified in the design step prior to gathering the data.

- **If $P \leq \alpha$, then**
  - the **decision** is to **reject** the null hypothesis (and conclude in favor of the alternative).
  - the results are characterized as **statistically significant**
  - the **conclusion** is: "At the $[\alpha]$ level, there is significant statistical evidence that $[H_1]$, verbally."

- **If $P > \alpha$, then**
  - the **decision** is to **not reject** the null hypothesis (and **not** conclude in favor of the alternative).
  - the results are characterized as **not statistically significant**
  - the **conclusion** is: "At the $[\alpha]$ level, there is **not** significant statistical evidence that $[H_A]$, verbally."
This is the “hypothesis testing” way, introduced in 1928-33, when Jerzy Neyman and Egon Pearson extended Karl Pearson's 1900 “significance testing” methodology. Nowadays, the terms significance testing and hypothesis testing are used interchangeably, and the distinction between the two methods is lost.

The modern method, combining the best of the two historic methods, is, based on the comparison of \( P \) and \( \alpha \), to

- make the decision to either reject or not reject the null hypothesis after the fashion of Neyman and Pearson,
- characterize the results as statistically significant or not statistically significant after the fashion of Karl Pearson and R.A. Fisher,
- and state the conclusion, always including the P-value, in the form "There [is/is not] significant statistical evidence that \([H_A, \text{ verbally}] \) \((___ < P < ___)\).

Specifically:

- If \( P \leq \alpha \), then
  - the decision is to reject the null hypothesis (and conclude in favor of the alternative).
  - the results are characterized as statistically significant.
  - the conclusion is: "There is significant statistical evidence that \([H_A, \text{ verbally}] \) \((___ < P < ___)\).

- If \( P > \alpha \), then
  - the decision is to not reject the null hypothesis (and not conclude in favor of the alternative).
  - the results are not statistically significant.
  - the conclusion is: "There is not significant statistical evidence that \([H_A, \text{ verbally}] \) \((___ < P < ___)\).

Exercises on Interpretation of P-Value

1. A test resulted in a P-value of 0.02. (a) Does this mean that the null hypothesis should be rejected, or NOT rejected, at the 0.05 level? (b) At the 0.01 level? (Hint: The phrase "rejected at the 0.05 level" means "rejected at the \( \alpha = 0.05 \) level of significance". In other words, the phrase "rejected at the 0.05 level" means that the following happened. During the study, at the design step, the investigator chose \( \alpha = 0.05 \) for the level of significance. Then, at the number
crunching step (Step 5), the investigator made the decision to reject the null hypothesis.

2. If a null hypothesis were rejected at the 0.05 level, would it have been rejected
   a. at the 0.10 level?
   b. At the 0.01 level? (Note: The moral of the story here is that if the P-value is between 0.01 and 0.10, one should report the P-value either exactly (in the form \( P = a \)) or as a range with a BOTH a lower and upper bound (in the form \( b < P < c \)), and it is not sufficient to report simply \( P < c \) or \( P > a \). BOTH the upper and lower bound are needed.)

**Example B: The 1-Sample T-Test, using the 6-Step Method**

The mean DBH of 16 trees randomly selected from Region 2 was 18 cm with a standard deviation of 10 cm. Is there evidence that the mean DBH of trees in Region 2 is greater than 11 cm?

**The 6-Step Method of Significance Testing**

1. **Model**
   a. Let \( X_i \) be the DBH (cm) of the \( i \)th randomly sampled tree from Region 2.
   b. Let \( E(X_i) = \text{(pop mean)} = \mu \) be the mean DBH (cm) of all trees in Region 2.
   c. Assume
      i. \( X_i \) distributed Normally

2. **Hypotheses**

   \[ H_0: \mu = 11, \ \text{versus} \ \ H_1: \mu > 11 \]

3. **Test Statistic**

   \[ T_{n-1} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \]

   [Because the parameter of interest is the population mean, and the population variance is not (assumed to be) known]

4. **Design**

   \( \alpha = 0.05 \)

   \( n = 16 \)
5. Gather data and analyze (compute)
   a. Best estimate of population mean = (sample mean) = \( \bar{X} = 18 \)
   b. Best estimate of SE is
      \[
      \frac{s}{\sqrt{n}} = \frac{10}{\sqrt{16}} = \frac{10}{4} = 2.5
      \]
   c. Test Statistic:
      \[
      T_{obs} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{18 - 11}{2.5} = 2.8
      \]
   d. P-value: Since \( H_1 \) specifies >, and since the degrees of freedom are \((n - 1) = 16 - 1 = 15,\)
      \[
      P = P\{T_{n-1} > T_{obs}\} = P\{T_{15} > 2.8\}
      \]
      Upon inspection of a t-table we find that \( T_{obs} = 2.8 \) is between \( t_{15} = 2.602 \) and \( t_{15} = 2.947, \) i.e.,
      \[
      (t_{15} = 2.602) < (T_{obs} = 2.8) < (t_{15} = 2.947),
      \]
      i.e.,
      \[
      (P\{T_{15} < 2.602\} = 0.99) < (P\{T_{15} < 2.8\}) < (P\{T_{15} < 2.947\} = 0.995)
      \]
      Therefore
      \[
      (P\{T_{15} > 2.602\} = 0.01) < (P\{T_{15} > 2.8\} = P\text{-value}) < (P\{T_{15} > 2.947\} = 0.005)
      \]
      i.e.,
      \[
      0.01 > P > 0.005
      \]
      or
      \[
      0.005 < P < 0.01.
      \]
   e. Comparison of \( P \) and \( \alpha: \) From the Design, Step 4, \( \alpha = 0.05. \) From the sample, Step 5, \( 0.005 < P < 0.01. \) Therefore, \( P < \alpha, \) and
      - the decision is to reject the null hypothesis at level \( \alpha = 0.05 \) (and conclude in favor of the alternative),
• the results are characterized as statistically significant, and

6. Conclusion

There is statistically significant evidence that the mean DBH of all trees in Region 2 is greater than 11 cm \((0.005 < P < 0.01)\).

Significance Testing Exercises on the 6-Step Method

For the next two problems, use the 6-step method of significance testing. Unless stated otherwise, use the 0.05 level of significance.

1. The mean height of mature citrus trees of a certain variety is reported to be 14.8 feet with a standard deviation of 1.5 feet. The Florida Citrus Growers Association believes that this figure for the mean height is too high. To test this they randomly selected 15 of these trees, measured their heights, and found the average height to be 13.6 feet. Assume that the height of such trees is normally distributed with a standard deviation of 1.5 feet.

2. According to the breeder's brochure, the average birth weight of the guinea pigs they sell is 29.5 grams. The birth weights of 20 guinea pigs bought from that breeder were 30, 30, 26, 32, 30, 23, 29, 31, 36, 30, 25, 34, 32, 24, 28, 27, 38, 31, 34, 30 grams. Assume that guinea pig birth weights are normally distributed, and test the hypothesis that the average birth weight is 29.5 grams. (Note: The computations can be done in JMP by entering the sample into a data table and then executing the commands Analyze, Distribution, Test Mean.)

3. Using the data of the previous exercise, test the hypothesis that the standard deviation of the birth weight of all guinea pigs sold by the breeder is greater than 3.0 g. To do the calculations in JMP, use the commands Analyze, Distribution, Test Standard Deviation. On the other hand, it would be a good idea to test your ability to do it with a hand calculator.

4. Recall the citrus tree problem (Exercise 1). (a) Recall the P-value you computed when you did this as a significance test. Based on that P-value, would the results have been significant at the 0.10 level? (b) At the 0.05 level? (c) At the 0.01 level? (d) At the 0.001 level?

5. Reconsider the guinea pig problem (Exercise 2) about mean birth weight. Without redoing the problem as a hypothesis test, just by looking at the observed P-value, answer the following questions with complete sentences. (a) Are the results significant at the 0.05 level? (b) Would you reject \(H_0\) at the 0.01 level? (c) At any level less than 0.10?

6. Reconsider the guinea pig problem (Exercise 3) about the standard deviation of birth weight. Without redoing the problem as a hypothesis test, just by looking at the observed P-value, answer the following questions with complete sentences. (a) Are the results significant at the 0.05 level? (b) Would you reject \(H_0\) at the 0.01 level? (c) At any level less than 0.10?
References


*Department of Statistics*

*Send Suggestions or Comments to Golde Holtzman*

*Last updated: 10/4/2012*


*URL: ../STAT5605/sigtest.pdf*