Chapter 7

Tests of Statistical Hypotheses

1 Tests about proportions

Example 1.1 Suppose a manufacturer of a certain printed circuit observes that about $p = 0.06$ of the circuits fail. An engineer and statistician working together suggest some changes that might improve the design of the product. To test this new procedure, it was agreed that $n = 200$ circuits would be produced using the proposed method and then checked. Let Y equal the number of these 200 circuits that fail. Clearly, if the number of failures, Y, is such that $Y/200$ is about equal to 0.06, then it seems that the new procedure has not resulted in an improvement. If Y is small so that $Y/200$ is about 0.02 or 0.03, we might believe that the new method is better than the old. On the other hand, if $Y/200$ is 0.09 or 0.10, the proposed method has perhaps caused a greater proportion of failures. What we need to establish is a formal rule that tells us when to accept the new procedure as an improvement. For example, we could accept the new procedure if $Y \leq 7$ or $Y/n \leq 0.035$. If we believe these trials, using the new procedure, are independent and have about the same probability of failure on each trial (i.e. i.i.d.), then $Y \sim B(200, p)$. Our goal is to make a statistical inference about p using the unbiased estimator $\hat{p} = Y/200$.

Definition 1.1 *Statistical hypothesis* is an assertion or conjecture about a population parameter. If the statistical hypothesis completely specifies the distribution, then it is call *simple*, otherwise, it is called *composite*.

Definition 1.2 Two complementary hypothesis in a statistical hypothesis testing problem are called the *null hypothesis* (denoted by H_0) and *alternative hypothesis* (denoted by H_1).

• In Example 1.1, The no change hypothesis $H_0 : p = 0.06$ is a simple null hypothesis. The hypothesis $H_1 : p < 0.06$ is a composite alternative hypothesis.

Definition 1.3 A *test of a statistical hypothesis* is a rule or procedure for deciding whether to reject H_0 and accept H_1 as true or to accept H_0 and reject H_1. We always say that H_0 is tested versus H_1.

• In Example 1.1, a possible test is: Reject $H_0 : p = 0.06$ and accept $H_1 : p < 0.06$ if $Y \leq 7$. The set $\{y : y \leq 7\}$ is called the *critical region.*

• Note that tests of $H_0 : p = 0.06$ versus $H_1 : p < 0.06$ or $H_0 : p = 0.06$ versus $H_1 : p > 0.06$ are called *one-sided tests.* Test of $H_0 : p = 0.06$ versus $H_1 : p \neq 0.06$ is called a *two-sided test.*

Types of errors and sizes of errors:

(i) Type I error: Reject H_0 when it is true.

(ii) Type II error: AcceptH_0 when it is false.

(iii) Size of a type I error=P(reject $H_0 \,|\, H_0$ is true)

(iii) Size of a type II error=P(accept $H_0 \,|\, H_0$ is false) $= \beta$

Definition 1.4 If H_0 is a simple hypothesis, then the *significance level of the test* is defined to be the size of the type I error (usually denoted by α). That is,

$$\alpha = \mathrm{P}(\text{reject } H_0 \,|\, H_0 \text{ is true}).$$

Example 1.2 In Example 1.1,

$$\alpha = \mathrm{P}(Y \leq 7 \,|\, p = 0.06) = \sum_{y=0}^{7} \binom{200}{y} (0.06)^y (0.94)^{200-y}.$$

Since $n = 200$ is large and p is small, α can be approximated by a Poisson distribution with $\lambda = np = 200(0.06) = 12$; i.e.,

$$\alpha \approx \sum_{y=0}^{7} \frac{12^y e^{-12}}{y!} = 0.09.$$

Suppose p has been improved to 0.03 (i.e. $H_1 : p = 0.03$), then

$$\begin{aligned}
\beta &= \mathrm{P}(Y > 7 \,|\, p = 0.03) = \sum_{y=8}^{200} \binom{200}{y} (0.03)^y (0.97)^{200-y} \\
&= 1 - \sum_{y=0}^{7} \binom{200}{y} (0.03)^y (0.97)^{200-y} \approx 1 - \sum_{y=0}^{7} \frac{6^y e^{-6}}{y!} = 0.256.
\end{aligned}$$

Remark 1.1 For a fixed sample size, it is impossible to make both types of error probabilities arbitrary small. In searching for a good test, it is common to restrict consideration to tests that control the type I error probability at a specified level. Within this class of tests, we search for tests that have Type II error probability that is as small as possible.

Tests hypothesis for one proportion:
Let Y be the number of successes in n independent Bernoulli trials with probability of success p. Then $Y \sim B(n,p)$. Consider testing $H_0 : p = p_0$ versus $H_1 : p > p_0$, where p_0 is some specified probability of success. By Central Limit Theorem, we know that when H_0 is true

$$\frac{Y/n - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1), \;\; n \to \infty,$$

or $Y/n \sim N(p_0, p_0(1-p_0)/n)$ approximately as n is large. Thus a test of $H_0 : p = p_0$ versus $H_1 : p > p_0$ is given by

$$\text{Reject } H_0 \text{ if and only if } Z = \frac{Y/n - p_0}{\sqrt{p_0(1-p_0)/n}} \geq z_\alpha \text{ or } Y/n \geq p_0 + z_\alpha \sqrt{p_0(1-p_0)/n}.$$

The significance level of this test is approximately α.

Example 1.3 It was claimed that many commercially manufactured dice are not fair because the "spots" are really indentations so that, for example, the 6-side is lighter than the 1-side. Let p equal the probability of rolling a 6 with one of these dice. To test $H_0 : p = 1/6$ against $H_1 : p > 1/6$, several of these dice will be rolled to yield a total of $n = 8000$ observations. Let Y be the number of times that six resulted in the 8000 trials. The test statistic is

$$Z = \frac{Y/8000 - 1/6}{\sqrt{(1/6)(5/6)/8000}}.$$

At the significance level of $\alpha = 0.05$, the critical region is

$$z \geq z_{0.05} = 1.645.$$

Suppose the results of the experiments yielded $y = 1389$. We have

$$z = \frac{1389/8000 - 1/6}{\sqrt{(1/6)(5/6)/8000}} = 1.670 > z_{0.05} = 1.645.$$

Thus H_0 is rejected and these experimental results indicate that these dice favor a 6 more than a fair die would.

• We summarize the tests of hypotheses for one proportion in Table 1.

Table 1: Tests of Hypotheses for One Proportion

H_0	H_1	Test Statistic	Critical Region
$p = p_0$	$p > p_0$		$Z \geq z_\alpha$
$p = p_0$	$p < p_0$	$Z = \dfrac{Y/n - p_0}{\sqrt{p_0(1-p_0)/n}}$	$Z \leq -z_\alpha$
$p = p_0$	$p \neq p_0$		$\|Z\| \geq z_{\alpha/2}$

• ***p*-value:** The *p-value* is the probability, under the null hypothesis H_0, that the test statistic is equal to or exceeds the observed value of the test statistic in the direction of the alternative hypothesis. The p-value is can be used in making a decision for a test. In Example 1.3, the observed value of the test statistic was $z = 1.67$. Since the alternative hypothesis was $H_1 : p > 1/6$, the p-value is

$$p\text{-value} = \mathrm{P}(Z \geq 1.67 \,|\, H_0) = 0.0475.$$

Since this p-value is less than $\alpha = 0.05$ and this would lead to rejection of H_0 at an $\alpha = 0.05$ significance level. If the alternative hypothesis is two-sided, i.e. $H_1 : p \neq 1/6$, the the p-value is $\mathrm{P}(|Z| \geq 1.67 \,|\, H_0) = 2\mathrm{P}(Z \geq 1.67 \,|\, H_0) = 0.095$ and would not lead to rejection of H_0 at $\alpha = 0.05$.

Tests hypothesis for two proportions:

Often there is interest in tests about p_1 and p_2, the probabilities of success for two different distributions. Let Y_1 and Y_2 be the number of successes in n_1 and n_2 independent trials with probabilities of success p_1 and p_2, respectively. Then $Y_1/n_1 \sim N(p_1, p_1(1-p_1)/n_1)$ approximately and $Y_2/n_2 \sim N(p_2, p_2(1-p_2)/n_2)$. Thus $Y_1/n_1 - Y_2/n_2 \sim N(p_1 - p_2, p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2)$ approximately. So

$$Z = \frac{Y_1/n_1 - Y_2/n_2 - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} \sim N(0,1) \text{ approximately.}$$

Consider testing $H_0 : p_1 = p_2 = p$. If p is unknown, we shall estimate p with $\hat{p} = (Y_1 + Y_2)/(n_1 + n_2)$. When H_0 is true, we obtain the test statistic

$$Z = \frac{Y_1/n_1 - Y_2/n_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$$

which has an approximate $N(0,1)$ distribution. Three tests of hypotheses for two proportions are summarized in Table 2

Table 2: Tests of Hypotheses for Two Proportions

H_0	H_1	Test Statistic	Critical Region
$p_1 = p_2$	$p_1 > p_2$		$Z \geq z_\alpha$
$p_1 = p_2$	$p_1 < p_2$	$Z = \frac{Y_1/n_1 - Y_2/n_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$	$Z \leq -z_\alpha$
$p_1 = p_2$	$p_1 \neq p_2$		$\lvert Z\rvert \geq z_{\alpha/2}$

Example 1.4 (Exercise 7.1-18, p352) The April 18, 1994, issue of *Time* magazine reported the results of a telephone poll of 800 adult Americans, 605 of them nonsmokers, who were asked the following question: "Should the federal tax on cigarettes be raised by \$1.25 to pay for health reform?" Let p_1 and p_2 equal the proportions of nonsmokers, respectively, who would say yes to this question. Given that $y_1 = 351$ nonsmokers and $y_2 = 41$ smokers said yes. (a) With $\alpha = 0.05$, test $H_0 : p_1 = p_2$ against $H_1 : p_1 \neq p_2$. (b) Find a 95% confidence interval for $p_1 - p_2$. Is your interval in agreement with the conclusion of part (a)? (c) Find a 95% confidence interval for p, the proportion of adult Americans who would say yes.

2 Tests about one mean and one variance

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. In this section, we study the tests about the mean μ and variance σ^2.

(i) Tests of $H_0 : \mu = \mu_0$ when σ^2 is known.

Consider testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$. We know that under H_0

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1).$$

When the sample mean $\overline{X}$ is close to μ_0 (i.e. Z is small), we tend to accept H_0. However, if $\overline{X}$ is far away from μ_0 (i.e. Z is large), we tend to reject H_0. Thus a possible test is given by

$$\text{Reject } H_0 \text{ if and only if } Z \geq z_\alpha \text{ or } \overline{X} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Tests of $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$ and $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ can be obtained in a similar fashion. These tests are summarized in Table 3.

Table 3: Tests of Hypotheses about μ, σ^2 known.

H_0	H_1	Test Statistic	Critical Region
$\mu = \mu_0$	$\mu > \mu_0$		$Z \geq z_\alpha$ or $\overline{X} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$
$\mu = \mu_0$	$\mu < \mu_0$	$Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$	$Z \leq -z_\alpha$ or $\overline{X} \leq \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$
$\mu = \mu_0$	$\mu \neq \mu_0$		$\lvert Z\rvert \geq z_{\alpha/2}$ or $\lvert\overline{X} - \mu_0\rvert \geq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Example 2.1 Let $X_1, \ldots, X_{52}$ is a random sample from $N(\mu, 100)$. Consider testing $H_0 : \mu = 60$ versus $H_1 : \mu > 60$. Suppose we obtain the observed sample mean $\overline{x} = 62.75$. For a significance level of $\alpha = 0.05$, we see that $\overline{x} = 62.75 > 62.718 = 60 + (1.645)\left(\frac{10}{\sqrt{52}}\right)$. Hence we tend to reject H_0. Note that $\overline{X} \sim N(60, 100/52)$ under H_0 and

$$\begin{aligned}
p\text{-value} &= \mathrm{P}(\overline{X} \geq 62.75 \mid \mu = 60) \\
&= \mathrm{P}\left(\frac{\overline{X} - 60}{10/\sqrt{52}} \geq \frac{62.75 - 60}{10/\sqrt{52}} \mid \mu = 60\right) \\
&= 1 - \Phi\left(\frac{62.75 - 60}{10/\sqrt{52}}\right) = 1 - \Phi(1.983) = 0.0237.
\end{aligned}$$

Since the p-value is less than $\alpha = 0.05$, we get the same decision of rejection. Note that if the alternative is two-sided $H_1 : \mu \neq 60$, then p-value$= 2(0.0237) = 0.0474$.

(ii) Tests of $H_0 : \mu = \mu_0$ when σ^2 is unknown.

When σ^2 is unknown, we know that under H_0

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1).$$

Therefore tests about $H_0 : \mu = \mu_0$ can be obtained in a similar way as case (i). These tests are summarized in Table 4.

Table 4: Tests of Hypotheses about μ, σ^2 unknown.

H_0	H_1	Test Statistic	Critical Region
$\mu = \mu_0$	$\mu > \mu_0$		$T \geq t_\alpha(n-1)$ or $\overline{X} \geq \mu_0 + t_\alpha(n-1)\frac{S}{\sqrt{n}}$
$\mu = \mu_0$	$\mu < \mu_0$	$T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$T \leq -t_\alpha(n-1)$ or $\overline{X} \leq \mu_0 - t_\alpha(n-1)\frac{S}{\sqrt{n}}$
$\mu = \mu_0$	$\mu \neq \mu_0$		$\lvert T\rvert \geq t_{\alpha/2}(n-1)$ or $\lvert\overline{X} - \mu_0\rvert \geq t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}$

Example 2.2 Let X (in millimeters) be the growth in 15 days of a tumor induced in a mouse. Assume that the distribution of X is $N(\mu, \sigma^2)$. To test $H_0 : \mu = 4.0$ against $H_1 : \mu \neq 4.0$, we take a random sample of size $n = 9$ and observe that $\overline{x} = 4.3$ and $s = 1.2$. Do we accept or reject H_0 at the 10% significance level?

Solution: Since

$$|t| = \frac{|\overline{x} - 4.0|}{s/\sqrt{9}} = \frac{|4.3 - 4.0|}{1.2/\sqrt{9}} = |0.75| < t_{0.05}(8) = 1.860,$$

we do not reject H_0 at the 10% significance level. Note that

$$p\text{-value} = P(|T| \geq 0.75 \,|\, H_0) = 2P(T \geq 0.75 \,|\, H_0).$$

From the table, we see that $0.706 = t_{0.25}(8) < 0.75 < 1.397 = t_{0.10}(8)$, so $0.10 < P(T \geq 0.75 \,|\, H_0) < 0.25$. Hence $0.20 < p$-value < 0.50 and this gives the same result. The computer gives the p-value to be 0.475.

(iii) Paired t tests:

Sometimes we are interesting in comparing the means of two different distributions or populations. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be n pairs of dependent measurements and let $D_i = X_i - Y_i,\ \imath = 1, 2, \ldots, n$. Assume $D_1, \ldots, D_n$ is a random sample from $N(\mu_D, \sigma^2)$. Consider the null hypothesis $H_0 : \mu_X = \mu_Y$, or equivalently, $H_0 : \mu_D = 0$. Then under H_0

$$T = \frac{\overline{D} - 0}{S_D/\sqrt{n}} \sim t(n-1).$$

Hence the appropriate t-test for a single mean could be used, selecting from Table 4.

Example 2.3 Twenty-four girls in the 9th and 10th grades were put on an ultra-heavy rope jumping program. Someone thought that such a program would increase their speed when running the 20-yard dash. Let D equal the difference in time to run the 40-yard dash-the "before program time" minus the "after program time." Assume that $D \sim N(\mu_D, \sigma_D^2)$ (approximately). We shall test $H_0 : \mu_D = 0$ against $H_1 : \mu_D > 0$ at a significance level of $\alpha = 0.05$. Suppose 24 observations of D are used and from these data we get $\overline{d} = 0.079$ and $s_d = 0.255$. Then the test statistic is given by

$$t = \frac{0.079 - 0}{0.255/\sqrt{24}} = 1.518 < t_{0.05}(23) = 1.714.$$

Thus we do not reject H_0. Note that

$$p\text{-value} = P(T \geq 1.518 \,|\, H_0).$$

From the table, we see that $1.319 = t_{0.10}(23) < 1.518 < 1.714 = t_{0.05}(23)$. Hence $0.05 <$ p-value < 0.10. The p-value also gives the same conclusion.

(iv) Tests of hypotheses about σ^2, μ unknown:

Consider testing $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$. When μ is unknown, we know that under H_0

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1).$$

Clearly, when S^2 is large (i.e. $\frac{(n-1)S^2}{\sigma_0^2}$ is large), we tend to reject H_0. Thus a possible test is given by

$$\text{Reject } H_0 \text{ if and only if } \frac{(n-1)S^2}{\sigma_0^2} \geq \chi_\alpha^2(n-1) \text{ or } S^2 \geq \frac{\sigma_0^2 \chi_\alpha^2(n-1)}{n-1}.$$

Tests about $H_0 : \sigma^2 = \sigma_0^2$ are summarized in Table 5.

Table 5: Tests of Hypotheses about σ^2, μ unknown.

H_0	H_1	Test Statistic	Critical Region
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$		$\frac{(n-1)S^2}{\sigma_0^2} \geq \chi^2_\alpha(n-1)$ or $S^2 \geq \frac{\sigma_0^2 \chi^2_\alpha(n-1)}{n-1}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\dfrac{(n-1)S^2}{\sigma_0^2}$	$\frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{1-\alpha}(n-1)$ or $S^2 \leq \frac{\sigma_0^2 \chi^2_{1-\alpha}(n-1)}{n-1}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$		$\frac{(n-1)S^2}{\sigma_0^2} \geq \chi^2_{\alpha/2}(n-1)$ or $\frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{1-\alpha/2}(n-1)$
			or $S^2 \geq \frac{\sigma_0^2 \chi^2_{\alpha/2}(n-1)}{n-1}$ or $S^2 \leq \frac{\sigma_0^2 \chi^2_{1-\alpha/2}(n-1)}{n-1}$

Example 2.4 A psychology professor claims that the variance of IQ scores for college students is equal to $\sigma^2 = 100$. To test this claim, it is decided to test the hypothesis $H_0 : \sigma^2 = 100$ against $H_0 : \sigma^2 \neq 100$. Suppose a random sample of $n = 23$ students is selected and $s^2 = 147.82$. At $\alpha = 0.05$ significance level, H_0 will be rejected if

$$\frac{(n-1)s^2}{\sigma_0^2} = \frac{(22)(147.82)}{100} = 32.52 \geq \chi^2_{\alpha/2}(22) = \chi^2_{0.025}(22)$$
$$\text{or } \frac{(n-1)s^2}{\sigma_0^2} = 32.52 \leq \chi^2_{1-\alpha/2}(22) = \chi^2_{0.975}(22).$$

Since $\chi^2_{0.025}(22) = 36.78 > 32.52 > \chi^2_{0.975}(22) = 10.98$, we do not reject H_0.

(v)Tests of hypotheses about σ^2, μ known:

When μ is known, we know that under H_0

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma_0^2} \sim \chi^2(n).$$

Thus tests about $H_0 : \sigma^2 = \sigma_0^2$ can be obtained in a similar way as case (iv). These tests are summarized in Table 6.

Table 6: Tests of Hypotheses about σ^2, μ known.

H_0	H_1	Test Statistic	Critical Region
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$		$\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma_0^2} \geq \chi^2_\alpha(n)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\dfrac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma_0^2}$	$\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma_0^2} \leq \chi^2_{1-\alpha}(n)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$		$\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma_0^2} \geq \chi^2_{\alpha/2}(n)$
			or $\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma_0^2} \leq \chi^2_{1-\alpha/2}(n)$

3 Tests of the equality of two normal distributions

Let independent random variables X and Y have normal distributions $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$, respectively.

(i) Tests of hypotheses for the equality of two means:

Assume that $\sigma_X^2 = \sigma_Y^2$ (unknown). Consider a test of $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X > \mu_Y$. Let

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{1/n + 1/m}},$$

where

$$S_p = \sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}.$$

Then $T \sim t(n+m-2)$ under H_0. A possible test is then given by

$$\text{Reject } H_0 \text{ if and only if } T \geq t_\alpha(n+m-2).$$

Tests about $H_0 : \mu_X = \mu_Y$ are summarized in Table 7.

Table 7: Tests of Hypotheses for equality of two means.

H_0	H_1	Test Statistic	Critical Region
$\mu_X = \mu_Y$	$\mu_X > \mu_Y$		$T \geq t_\alpha(n+m-2)$
$\mu_X = \mu_Y$	$\mu_X < \mu_Y$	$T = \frac{\overline{X}-\overline{Y}}{S_p\sqrt{1/n+1/m}}$	$T \leq -t_\alpha(n+m-2)$
$\mu_X = \mu_Y$	$\mu_X \neq \mu_Y$		$\lvert T\rvert \geq t_{\alpha/2}(n+m-2)$

Example 3.1 A product is packaged using a machine with 24 filler heads numbered 1 to 24, with the odd numbered heads on one side of machine and the even on the other side. Let X and Y equal the fill weights in grams when a package is filled by an odd-numbered head and an even-numbered head, respectively. Assume that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ and X and Y are independent. Would like to test $H_0 : \mu_X = \mu_Y$ against $\mu_X \neq \mu_Y$. To perform the test, after the machine has been set up and is running, we select one package at random from each filler head and weight it. Take $n = m = 12$ and from the data we get $\overline{x} = 1076.75$, $s_x^2 = 29.30$, $\overline{y} = 1072.33$ and $s_y^2 = 26.24$. Then the

value of the test statistic is

$$t = \frac{1076.75 - 1072.33}{\sqrt{\frac{11(29.30)+11(26.24)}{22}\left(\frac{1}{12}+\frac{1}{12}\right)}} = 2.05.$$

At an $\alpha = 0.10$ significance level, we have

$$|t| = 2.05 > t_{0.05}(22) = 1.717.$$

We conclude that H_0 is rejected at $\alpha = 0.10$ significance level. Note that

$$|t| = 2.05 < t_{0.025}(22) = 2.074.$$

So H_0 would not be rejected at $\alpha = 0.05$ significance level. That is the p-value is between 0.05 and 0.10.

Discussions: If the variances of X and Y are known, then the statistic for testing $H_0 : \mu_X = \mu_Y$ is

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}},$$

and $Z \sim N(0,1)$ under H_0. If the variances are unknown and the sample sizes are large, the appropriate test statistic is

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}},$$

and $Z \sim N(0,1)$ approximately.

Example 3.2 The target thickness for Fruit Flavored Gum and for Fruit Flavored Bubble Gum is 6.7 hundredths of an inch. Let the independent random variables X and Y equal the respective thickness of these gums in hundredths of an inch and assume that their distributions are $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$, respectively. Because bubble gum has more elasticity than regular gum, it seems as if it would be harder to roll it out to the correct thickness. Thus we shall test $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$ using sample sizes $n = 50$ and $m = 40$. Suppose from the data we have $\overline{x} = 6.701$, $s_x = 0.108$, $\overline{y} = 6.841$ and $s_y = 0.155$. Since at an approximate significance level of $\alpha = 0.01$ the test statistic is

$$z = \frac{6.701 - 6.841}{\sqrt{\frac{0.108^2}{50} + \frac{0.155^2}{40}}} = -4.848 < -2.326 = -z_{0.01},$$

the null hypothesis is clearly rejected.

(ii) Tests of hypotheses for the equality of two variances:

Consider testing $H_0 : \sigma_X^2 = \sigma_Y^2$ (or $\sigma_X^2/\sigma_Y^2 = 1$). Take random samples of n observations of X and m observations of Y. Recall that $(n-1)S_X^2/\sigma_X^2 \sim \chi^2(n-1)$ and $(m-1)S_Y^2/\sigma_Y^2 \sim \chi^2(m-1)$. Thus when H_0 is true,

$$F = \frac{\frac{(n-1)S_X^2}{\sigma_X^2}/(n-1)}{\frac{(m-1)S_Y^2}{\sigma_Y^2}/(m-1)} = \frac{S_X^2}{S_Y^2} \sim F(n-1, m-1).$$

When H_0 is true, we would expect the observed value of F to be close to 1. Tests about $H_0 : \sigma_X^2 = \sigma_Y^2$ are summarized in Table 8.

Table 8: Tests of Hypotheses for equality of variances.

H_0	H_1	Test Statistic	Critical Region
$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 > \sigma_Y^2$		$F \geq F_\alpha(n-1, m-1)$
$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 < \sigma_Y^2$	$F = \frac{S_X^2}{S_Y^2}$	$F \leq F_{1-\alpha}(n-1, m-1)$
$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 \neq \sigma_Y^2$		$F \geq F_{\alpha/2}(n-1, m-1)$ or $F \leq F_{1-\alpha/2}(n-1, m-1)$

Remark 3.1 $F_{1-\alpha}(n-1, m-1) = \dfrac{1}{F_\alpha(m-1, n-1)}$.

Example 3.3 To measure air pollution in a home, let X and Y equal the amount of suspended particulate matter (in μg/m^3) measured during a 24-hour period in a home in which there is no smoker and a home in which there is a smoker, respectively. (Assume that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$.) We shall test the null hypothesis $H_0 : \sigma_X^2/\sigma_Y^2 = 1$ against the alternative hypothesis $H_1 : \sigma_X^2/\sigma_Y^2 > 1$. If a random sample of size $n = 31$ yielded $\overline{x} = 93$ and $s_x = 12.9$ while a random sample of size $m = 31$ yielded $\overline{y} = 132$ and $s_y = 7.1$. Define a critical region and give your conclusion if $\alpha = 0.05$. Now test $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$ if $\alpha = 0.05$.