CHAPTER 10

Statistical Decision Theory

STATISTICAL DECISIONS

Very often in practice we are called upon to make decisions about populations on the basis of sample information. Such decisions are called *statistical decisions*. For example, we may wish to decide on the basis of sample data whether a new serum is really effective in curing a disease, whether one educational procedure is better than another, or whether a given coin is loaded.

STATISTICAL HYPOTHESES

In attempting to reach decisions, it is useful to make assumptions (or guesses) about the populations involved. Such assumptions, which may or may not be true, are called *statistical hypotheses*. They are generally statements about the probability distributions of the populations.

Null Hypotheses

In many instances we formulate a statistical hypothesis for the sole purpose of rejecting or nullifying it. For example, if we want to decide whether a given coin is loaded, we formulate the hypothesis that the coin is fair (i.e., $p = 0.5$, where p is the probability of heads). Similarly, if we want to decide whether one procedure is better than another, we formulate the hypothesis that there is *no difference* between the procedures (i.e., any observed differences are due merely to fluctuations in sampling from the *same* population). Such hypotheses are often called *null hypotheses* and are denoted by H_0.

Alternative Hypotheses

Any hypothesis that differs from a given hypothesis is called an *alternative hypothesis*. For example, if one hypothesis is $p = 0.5$, alternative hypotheses might be $p = 0.7$, $p \neq 0.5$, or $p > 0.5$. A hypothesis alternative to the null hypothesis is denoted by H_1.

TESTS OF HYPOTHESES AND SIGNIFICANCE, OR DECISION RULES

If we suppose that a particular hypothesis is true but find that the results observed in a random sample differ markedly from the results expected under the hypothesis (i.e., expected on the basis of pure chance, using sampling theory), then we would say that the observed differences are *significant* and would thus be inclined to reject the hypothesis (or at least not accept it on the basis of the evidence obtained). For example, if 20 tosses of a coin yield 16 heads, we would be inclined to reject the hypothesis that the coin is fair, although it is conceivable that we might be wrong.

Procedures that enable us to determine whether observed samples differ significantly from the results expected, and thus help us decide whether to accept or reject hypotheses, are called *tests of hypotheses, tests of significance, rules of decision*, or simply *decision rules*.

TYPE I AND TYPE II ERRORS

If we reject a hypothesis when it should be accepted, we say that a *Type I error* has been made. If, on the other hand, we accept a hypothesis when it should be rejected, we say that a *Type II error* has been made. In either case, a wrong decision or error in judgment has occurred.

In order for decision rules (or tests of hypotheses) to be good, they must be designed so as to minimize errors of decision. This is not a simple matter, because for any given sample size, an attempt to decrease one type of error is generally accompanied by an increase in the other type of error. In practice, one type of error may be more serious than the other, and so a compromise should be reached in favor of limiting the more serious error. The only way to reduce both types of error is to increase the sample size, which may or may not be possible.

LEVEL OF SIGNIFICANCE

In testing a given hypothesis, the maximum probability with which we would be willing to risk a Type I error is called the *level of significance*, or *significance level*, of the test. This probability, often denoted by α, is generally specified before any samples are drawn so that the results obtained will not influence our choice.

In practice, a significance level of 0.05 or 0.01 is customary, although other values are used. If, for example, the 0.05 (or 5%) significance level is chosen in designing a decision rule, then there are about 5 chances in 100 that we would reject the hypothesis when it should be accepted; that is, we are about 95% *confident* that we have made the right decision. In such case we say that the hypothesis has been rejected at the 0.05 significance level, which means that the hypothesis has a 0.05 probability of being wrong.

TESTS INVOLVING NORMAL DISTRIBUTIONS

To illustrate the ideas presented above, suppose that under a given hypothesis the sampling distribution of a statistic S is a normal distribution with mean μ_S and standard deviation σ_S. Thus the distribution of the standardized variable (or z score), given by $z = (S - \mu_S)/\sigma_S$, is the standardized normal distribution (mean 0, variance 1), as shown in Fig. 10-1.

As indicated in Fig. 10-1, we can be 95% confident that if the hypothesis is true, then the z score of an actual sample statistic S will lie between -1.96 and 1.96 (since the area under the normal curve between these values is 0.95). However, if on choosing a single sample at random we find that the z score of its statistic lies *outside* the range -1.96 to 1.96, we would conclude that such an event could happen with a probability of only 0.05 (the total shaded area in the figure) if the given hypothesis were true. We would then say that this z score differed *significantly* from what would be expected under the hypothesis, and we would then be inclined to reject the hypothesis.

The total shaded area 0.05 is the significance level of the test. It represents the probability of our being wrong in rejecting the hypothesis (i.e., the probability of making a Type I error). Thus we say that

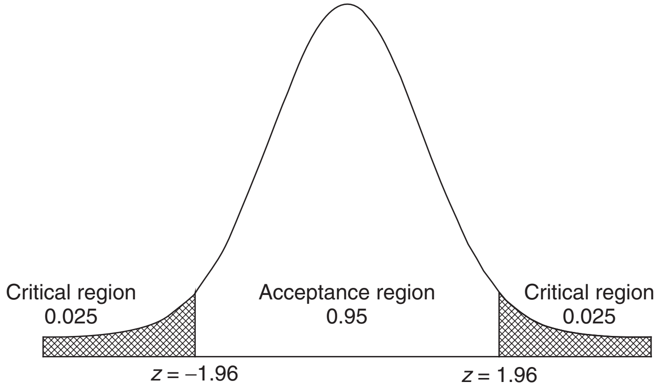


Fig. 10-1 Standard normal curve with the critical region (0.05) and acceptance region (0.95).

the hypothesis is *rejected at 0.05 the significance level* or that the z score of the given sample statistic is *significant at the 0.05 level*.

The set of z scores outside the range -1.96 to 1.96 constitutes what is called the *critical region of the hypothesis, the region of rejection of the hypothesis*, or *the region of significance*. The set of z scores inside the range -1.96 to 1.96 is thus called the *region of acceptance of the hypothesis*, or the *region of non-significance*.

On the basis of the above remarks, we can formulate the following decision rule (or test of hypothesis or significance):

> Reject the hypothesis at the 0.05 significance level if the z score of the statistic S lies outside the range -1.96 to 1.96 (i.e., either $z > 1.96$ or $z < -1.96$). This is equivalent to saying that the observed sample statistic is significant at the 0.05 level.
>
> Accept the hypothesis otherwise (or, if desired, make no decision at all).

Because the z score plays such an important part in tests of hypotheses, it is also called a *test statistic*.

It should be noted that other significance levels could have been used. For example, if the 0.01 level were used, we would replace 1.96 everywhere above with 2.58 (see Table 10.1). Table 9.1 can also be used, since the sum of the significance and confidence levels is 100%.

Table 10.1

Level of significance, α	0.10	0.05	0.01	0.005	0.002
Critical values of z for one-tailed tests	-1.28 *or* 1.28	-1.645 *or* 1.645	-2.33 *or* 2.33	-2.58 *or* 2.58	-2.88 *or* 2.88
Critical values of z for two-tailed tests	-1.645 *and* 1.645	-1.96 *and* 1.96	-2.58 *and* 2.58	-2.81 *and* 2.81	-3.08 *and* 3.08

TWO-TAILED AND ONE-TAILED TESTS

In the above test we were interested in extreme values of the statistic S or its corresponding z score on *both* sides of the mean (i.e., in both tails of the distribution). Such tests are thus called *two-sided tests*, or *two-tailed tests*.

Often, however, we may be interested only in extreme values to one side of the mean (i.e., in one tail of the distribution), such as when we are testing the hypothesis that one process is better than another (which is different from testing whether one process is better or worse than the other). Such tests are called *one-sided tests*, or *one-tailed tests*. In such cases the critical region is a region to one side of the distribution, with area equal to the level of significance.

Table 10.1, which gives critical values of z for both one-tailed and two-tailed tests at various levels of significance, will be found useful for reference purposes. Critical values of z for other levels of significance are found from the table of normal-curve areas (Appendix II).

SPECIAL TESTS

For large samples, the sampling distributions of many statistics are normal distributions (or at least nearly normal), and the above tests can be applied to the corresponding z scores. The following special cases, taken from Table 8.1, are just a few of the statistics of practical interest. In each case the results hold for infinite populations or for sampling with replacement. For sampling without replacement from finite populations, the results must be modified. See page 182.

1. **Means.** Here $S = \bar{X}$, the sample mean; $\mu_S = \mu_{\bar{X}} = \mu$, the population mean; and $\sigma_S = \sigma_{\bar{X}} = \sigma/\sqrt{N}$, where σ is the population standard deviation and N is the sample size. The z score is given by

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$

When necessary, the sample deviation s or $\hat{s}$ is used to estimate σ.

2. **Proportions.** Here $S = P$, the proportion of "successes" in a sample; $\mu_S = \mu_P = p$, where p is the population proportion of successes and N is the sample size; and $\sigma_S = \sigma_P = \sqrt{pq/N}$, where $q = 1 - p$.

The z score is given by

$$z = \frac{P - p}{\sqrt{pq/N}}$$

In case $P = X/N$, where X is the actual number of successes in a sample, the z score becomes

$$z = \frac{X - Np}{\sqrt{Npq}}$$

That is, $\mu_X = \mu = Np$, $\sigma_X = \sigma = \sqrt{Npq}$, and $S = X$.

The results for other statistics can be obtained similarly.

OPERATING-CHARACTERISTIC CURVES; THE POWER OF A TEST

We have seen how the Type I error can be limited by choosing the significance level properly. It is possible to avoid risking Type II errors altogether simply by not making them, which amounts to never accepting hypotheses. In many practical cases, however, this cannot be done. In such cases, use is often made of *operating-characteristic curves*, or *OC curves*, which are graphs showing the probabilities of Type II errors under various hypotheses. These provide indications of how well a given test will enable us to minimize Type II errors; that is, they indicate the *power of a test* to prevent us from making wrong decisions. They are useful in designing experiments because they show such things as what sample sizes to use.

p-VALUES FOR HYPOTHESES TESTS

The p-value is the probability of observing a sample statistic as extreme or more extreme than the one observed under the assumption that the null hypothesis is true. When testing a hypothesis, state the value of α. Calculate your p-value and if the p-value $\leq \alpha$, then reject H_0. Otherwise, do not reject H_0. For testing means, using large samples ($n > 30$), calculate the p-value as follows:

1. For $H_0: \mu = \mu_0$ and $H_1: \mu < \mu_0$, p-value $= P(Z <$ computed test statistic),
2. For $H_0: \mu = \mu_0$ and $H_1: \mu > \mu_0$, p-value $= P(Z >$ computed test statistic), and
3. For $H_0: \mu = \mu_0$ and $H_1: \mu \neq \mu_0$, p-value $= P(Z < -|$ computed test statistic $|) + P(Z > |$ computed test statistic $|)$.

The computed test statistic is $\dfrac{\bar{x} - \mu_0}{(s/\sqrt{n})}$, where $\bar{x}$ is the mean of the sample, s is the standard deviation of the sample, and μ_0 is the value specified for μ in the null hypothesis. Note that if σ is unknown, it is estimated from the sample by using s. This method of testing hypothesis is equivalent to the method of finding a critical value or values and if the computed test statistic falls in the rejection region, reject the null hypothesis. The same decision will be reached using either method.

CONTROL CHARTS

It is often important in practice to know when a process has changed sufficiently that steps should be taken to remedy the situation. Such problems arise, for example, in quality control. Quality control supervisors must often decide whether observed changes are due simply to chance fluctuations or are due to actual changes in a manufacturing process because of deteriorating machine parts, employees' mistakes, etc. *Control charts* provide a useful and simple method for dealing with such problems (see Problem 10.16).

TESTS INVOLVING SAMPLE DIFFERENCES

Differences of Means

Let $\bar{X}_1$ and $\bar{X}_2$ be the sample means obtained in large samples of sizes N_1 and N_2 drawn from respective populations having means μ_1 and μ_2 and standard deviations σ_1 and σ_2. Consider the null hypothesis that there is *no difference* between the population means (i.e., $\mu_1 = \mu_2$), which is to say that the samples are drawn from two populations having the same mean.

Placing $\mu_1 = \mu_2$ in equation (5) of Chapter 8, we see that the sampling distribution of differences in means is approximately normally distributed, with its mean and standard deviation given by

$$\mu_{\bar{X}1-\bar{X}2} = 0 \qquad \text{and} \qquad \sigma_{\bar{X}1-\bar{X}2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \tag{1}$$

where we can, if necessary, use the sample standard deviations s_1 and s_2 (or $\hat{s}_1$ and $\hat{s}_2$) as estimates of σ_1 and σ_2.

By using the standardized variable, or z score, given by

$$z = \frac{\bar{X}_1 - \bar{X}_2 - 0}{\sigma_{\bar{X}1-\bar{X}2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}1-\bar{X}2}} \tag{2}$$

we can test the null hypothesis against alternative hypotheses (or the significance of an observed difference) at an appropriate level of significance.

Differences of Proportions

Let P_1 and P_2 be the sample proportions obtained in large samples of sizes N_1 and N_2 drawn from respective populations having proportions p_1 and p_2. Consider the null hypothesis that there is *no difference* between the population parameters (i.e., $p_1 = p_2$) and thus that the samples are really drawn from the same population.

Placing $p_1 = p_2 = p$ in equation (6) of Chapter 8, we see that the sampling distribution of differences in proportions is approximately normally distributed, with its mean and standard deviation given by

$$\mu_{P1-P2} = 0 \qquad \text{and} \qquad \sigma_{P1-P2} = \sqrt{pq\left(\frac{1}{N_1} + \frac{1}{N_2}\right)} \tag{3}$$

where
$$p = \frac{N_1 P_1 + N_2 P_2}{N_1 + N_2}$$

is used as an estimate of the population proportion and where $q = 1 - p$.

By using the standardized variable

$$z = \frac{P_1 - P_2 - 0}{\sigma_{P1-P2}} = \frac{P_1 - P_2}{\sigma_{P1-P2}} \tag{4}$$

we can test observed differences at an appropriate level of significance and thereby test the null hypothesis.

Tests involving other statistics can be designed similarly.

TESTS INVOLVING BINOMIAL DISTRIBUTIONS

Tests involving binomial distributions (as well as other distributions) can be designed in a manner analogous to those using normal distributions; the basic principles are essentially the same. See Problems 10.23 to 10.28.

Solved Problems

TESTS OF MEANS AND PROPORTIONS, USING NORMAL DISTRIBUTIONS

10.1 Find the probability of getting between 40 and 60 heads inclusive in 100 tosses of a fair coin.

SOLUTION

According to the binomial distribution, the required probability is

$$\binom{100}{40}\left(\frac{1}{2}\right)^{40}\left(\frac{1}{2}\right)^{60} + \binom{100}{41}\left(\frac{1}{2}\right)^{41}\left(\frac{1}{2}\right)^{59} + \cdots + \binom{100}{60}\left(\frac{1}{2}\right)^{60}\left(\frac{1}{2}\right)^{40}$$

Since $Np = 100(\frac{1}{2})$ and $Nq = 100(\frac{1}{2})$ are both greater than 5, the normal approximation to the binomial distribution can be used in evaluating this sum. The mean and standard deviation of the number of heads in 100 tosses are given by

$$\mu = Np = 100(\tfrac{1}{2}) = 50 \qquad \text{and} \qquad \sigma = \sqrt{Npq} = \sqrt{(100)(\tfrac{1}{2})(\tfrac{1}{2})} = 5$$

On a continuous scale, between 40 and 60 heads inclusive is the same as between 39.5 and 60.5 heads. We thus have

$$39.5 \text{ in standard units} = \frac{39.5 - 50}{5} = -2.10 \qquad 60.5 \text{ in standard units} = \frac{60.5 - 50}{5} = 2.10$$

$$\begin{aligned}\text{Required probability} &= \text{area under normal curve between } z = -2.10 \text{ and } z = 2.10\\ &= 2(\text{area between } z = 0 \text{ and } z = 2.10) = 2(0.4821) = 0.9642\end{aligned}$$

10.2 To test the hypothesis that a coin is fair, adopt the following decision rule:

Accept the hypothesis if the number of heads in a single sample of 100 tosses is between 40 and 60 inclusive.

Reject the hypothesis otherwise.

(*a*) Find the probability of rejecting the hypothesis when it is actually correct.

(*b*) Graph the decision rule and the result of part (*a*).

(*c*) What conclusions would you draw if the sample of 100 tosses yielded 53 heads? And if it yielded 60 heads?

(*d*) Could you be wrong in your conclusions about part (*c*)? Explain.

SOLUTION

(*a*) From Problem 10.1, the probability of not getting between 40 and 60 heads inclusive if the coin is fair is $1 - 0.9642 = 0.0358$. Thus the probability of rejecting the hypothesis when it is correct is 0.0358.

(*b*) The decision rule is illustrated in Fig. 10-2, which shows the probability distribution of heads in 100 tosses of a fair coin. If a single sample of 100 tosses yields a z score between -2.10 and 2.10, we accept the hypothesis; otherwise, we reject the hypothesis and decide that the coin is not fair.

The error made in rejecting the hypothesis when it should be accepted is the *Type 1 error* of the decision rule; and the probability of making this error, equal to 0.0358 from part (*a*), is represented by the total shaded area of the figure. If a single sample of 100 tosses yields a number of heads whose z score (or z statistic) lies in the shaded regions, we would say that this z score differed *significantly* from what would be expected if the hypothesis were true. For this reason, the total shaded area (i.e., the probability of a Type I error) is called the *significance level* of the decision rule and equals 0.0358 in this case. Thus we speak of rejecting the hypothesis at the 0.0358 (or 3.58%) significance level.

(*c*) According to the decision rule, we would have to accept the hypothesis that the coin is fair in both cases. One might argue that if only one more head had been obtained, we would have rejected the hypothesis. This is what one must face when any sharp line of division is used in making decisions.

(*d*) Yes. We could accept the hypothesis when it actually should be rejected—as would be the case, for example, when the probability of heads is really 0.7 instead of 0.5. The error made in accepting the hypothesis when it should be rejected is the *Type II error* of the decision.

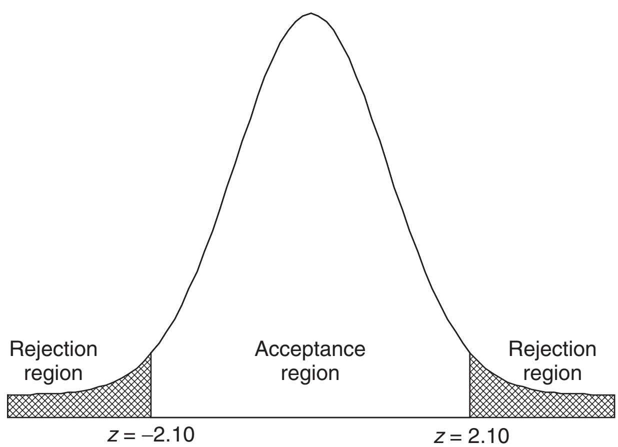


Fig. 10-2 Standard normal curve showing the acceptance and rejection regions for testing that the coin is fair.

10.3 Using the binomial distribution and not the normal approximation to the binomial distribution, design a decision rule to test the hypothesis that a coin is fair if a sample of 64 tosses of the coin is taken and a significance level of 0.05 is used. Use MINITAB to assist with the solution.

SOLUTION

The binomial plot of probabilities when a fair coin is tossed 64 times is given in Fig. 10.3. Partial cumulative probabilities generated by MINITAB are shown below the Fig. 10-3.

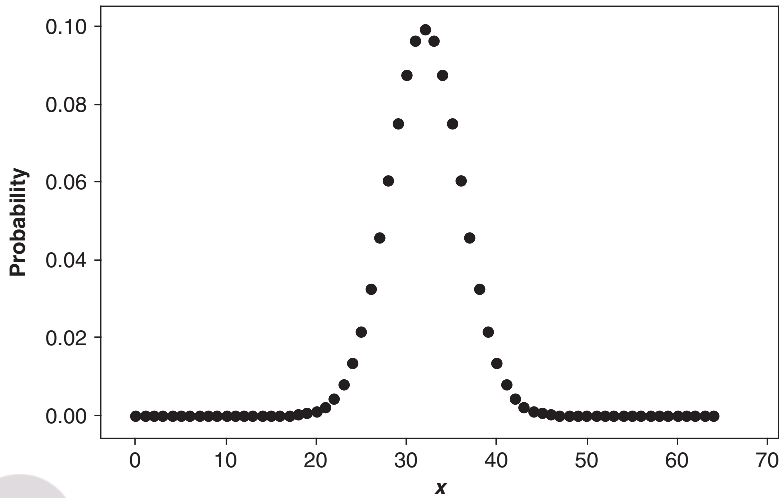


Fig. 10-3 MINITAB plot of the binomial distribution for $n = 64$ and $p = 0.5$.

x	Probability	Cumulative	x	Probability	Cumulative
0	0.0000000	0.0000000	13	0.0000007	0.0000009
1	0.0000000	0.0000000	14	0.0000026	0.0000035
2	0.0000000	0.0000000	15	0.0000086	0.0000122
3	0.0000000	0.0000000	16	0.0000265	0.0000387
4	0.0000000	0.0000000	17	0.0000748	0.0001134
5	0.0000000	0.0000000	18	0.0001952	0.0003087
6	0.0000000	0.0000000	19	0.0004727	0.0007814
7	0.0000000	0.0000000	20	0.0010636	0.0018450
8	0.0000000	0.0000000	21	0.0022285	0.0040735
9	0.0000000	0.0000000	22	0.0043556	0.0084291
10	0.0000000	0.0000000	23	0.0079538	0.0163829
11	0.0000000	0.0000001	24	0.0135877	0.0299706
12	0.0000002	0.0000002	25	0.0217403	0.0517109

We see that $P(X \leq 23) = 0.01638$. Because the distribution is symmetrical, we also know $P(X \geq 41) = 0.01638$. The rejection region $\{X \leq 23$ and $X \geq 41\}$ has probability $2(0.01638) = 0.03276$. The rejection region $\{X \leq 24$ and $X \geq 40\}$ would exceed 0.05. When the binomial distribution is used we cannot have a rejection region equal to exactly 0.05. The closest we can get to 0.05 without exceeding it is 0.03276.

Summarizing, the coin will be flipped 64 times. It will be declared unfair or not balanced if 23 or fewer heads or 41 or more heads are obtained. The chance of making a Type 1 error is 0.03276 which is as close to 0.05 as you can get without exceeding it.

10.4 Refer to Problem 10.3. Using the binomial distribution and not the normal approximation to the binomial distribution, design a decision rule to test the hypothesis that a coin is fair if a sample of 64 tosses of the coin is taken and a significance level of 0.05 is used. Use EXCEL to assist with the solution.

SOLUTION

The outcomes 0 through 64 are entered into column A of the EXCEL worksheet. The expressions =BINOMDIST(A1, 64, 0.5, 0) and =BINOMDIST(A1, 64, 0.5, 1) are used to obtain the binomial and cumulative binomial distributions. The 0 for the fourth parameter requests individual probabilities and the 1 requests cumulative probabilities. A click-and-drag in column B gives the individual probabilities, and in column C a click-and-drag gives the cumulative probabilities.

A	B	C	A	B	C
X	Probability	Cumulative	x	Probability	Cumulative
0	5.42101E-20	5.42101E-20	13	7.12151E-07	9.40481E-07
1	3.46945E-18	3.52366E-18	14	2.59426E-06	3.53474E-06
2	1.09288E-16	1.12811E-16	15	8.64754E-06	1.21823E-05
3	2.25861E-15	2.37142E-15	16	2.64831E-05	3.86654E-05
4	3.44438E-14	3.68152E-14	17	7.47758E-05	0.000113441
5	4.13326E-13	4.50141E-13	18	0.000195248	0.000308689
6	4.06437E-12	4.51451E-12	19	0.000472706	0.000781395
7	3.36762E-11	3.81907E-11	20	0.001063587	0.001844982
8	2.39943E-10	2.78134E-10	21	0.002228469	0.004073451
9	1.49298E-09	1.77111E-09	22	0.004355644	0.008429095
10	8.21138E-09	9.98249E-09	23	0.007953785	0.01638288
11	4.03104E-08	5.02929E-08	24	0.013587715	0.029970595
12	1.78038E-07	2.28331E-07	25	0.021740344	0.051710939

It is found, as in Problem 10.3, that $P(X \leq 23) = 0.01638$ and because of symmetry, $P(X \geq 41) = 0.01638$ and that the rejection region is $\{X \leq 23 \text{ or } X \geq 41\}$ and the significance level is $0.01638 + 0.01638$ or 0.03276.

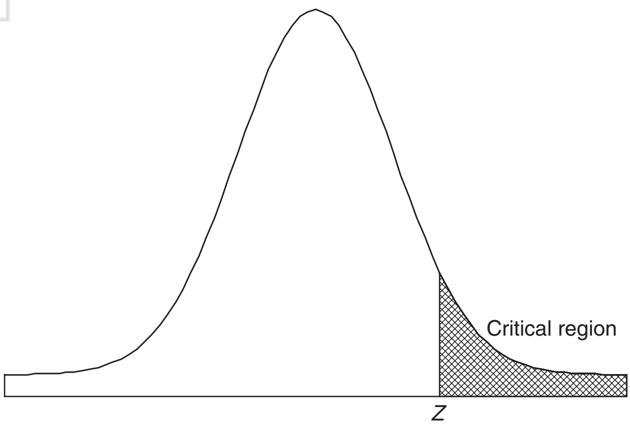


Fig. 10-4 Determining the Z value that will give a critical region equal to 0.05.

10.5 In an experiment on extrasensory perception (ESP), an individual (subject) in one room is asked to state the color (red or blue) of a card chosen from a deck of 50 well-shuffled cards by an individual in another room. It is unknown to the subject how many red or blue cards are in the deck. If the subject identifies 32 cards correctly, determine whether the results are significant at the (*a*) 0.05 and (*b*) 0.01 levels.

SOLUTION

If p is the probability of the subject choosing the color of a card correctly, then we have to decide between two hypotheses:

$H_0 : p = 0.5$, and the subject is simply guessing (i.e., the results are due to chance).

$H_1 : p > 0.5$, and the subject has powers of ESP.

Since we are not interested in the subject's ability to obtain extremely low scores, but only in the ability to obtain high scores, we choose a one-tailed test. If hypothesis H_0 is true, then the mean and standard deviation of the number of cards identified correctly are given by

$$\mu = Np = 50(0.5) = 25 \qquad \text{and} \qquad \sigma = \sqrt{Npq} = \sqrt{50(0.5)(0.5)} = \sqrt{12.5} = 3.54$$

(*a*) For a one-tailed test at the 0.05 significance level, we must choose z in Fig. 10-4 so that the shaded area in the critical region of high scores is 0.05. The area between 0 and z is 0.4500, and $z = 1.645$; this can also be read from Table 10.1. Thus our decision rule (or test of significance) is:

If the z score observed is greater than 1.645, the results are significant at the 0.05 level and the individual has powers of ESP.

If the z score is less than 1.645, the results are due to chance (i.e., not significant at the 0.05 level).

Since 32 in standard units is $(32 - 25)/3.54 = 1.98$, which is greater than 1.645, we conclude at the 0.05 level that the individual has powers of ESP.

Note that we should really apply a continuity correction, since 32 on a continuous scale is between 31.5 and 32.5. However, 31.5 has a standard score of $(31.5 - 25)/3.54 = 1.84$, and so the same conclusion is reached.

(*b*) If the significance level is 0.01, then the area between 0 and z is 0.4900, from which we conclude that $z = 2.33$.

Since 32 (or 31.5) in standard units is 1.98 (or 1.84), which is less than 2.33, we conclude that the results are *not significant* at the 0.01 level.

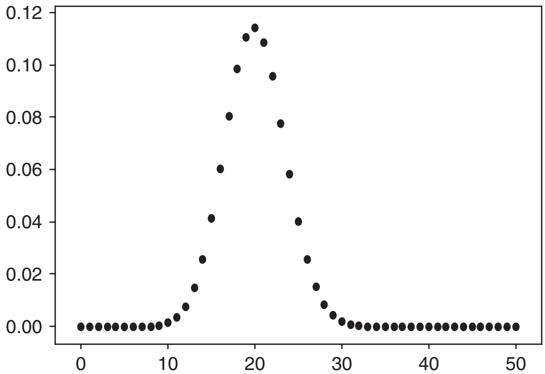
Some statisticians adopt the terminology that results significant at the 0.01 level are *highly significant*, that results significant at the 0.05 level but not at the 0.01 level are *probably significant*, and that results significant at levels larger than 0.05 are *not significant*. According to this terminology, we would conclude that the above experimental results are *probably significant*, so that further investigations of the phenomena are probably warranted.

Since significance levels serve as guides in making decisions, some statisticians quote the actual probabilities involved. For instance, since $pr\{z \geq 1.84\} = 0.0322$, in this problem, the statistician could say that on the basis of the experiment the chances of being wrong in concluding that the individual has powers of ESP are about 3 in 100. The quoted probability (0.0322 in this case) is called the *p*-value for the test.

10.6 The claim is made that 40% of tax filers use computer software to file their taxes. In a sample of 50, 14 used computer software to file their taxes. Test $H_0: p = 0.4$ versus $H_a: p < 0.4$ at $\alpha = 0.05$ where p is the population proportion who use computer software to file their taxes. Test using the binomial distribution and test using the normal approximation to the binomial distribution.

SOLUTION

If the exact test of $H_0: p = 0.4$ versus $H_a: p < 0.4$ at $\alpha = 0.05$ is used, the null is rejected if $X \leq 15$. This is called the rejection region. If the test based on the normal approximation to the binomial is used, the null is rejected if $Z < -1.645$ and this is called the rejection region. $X = 14$ is called the test statistic. The binomial test statistic is in the rejection region and the null is rejected. Using the normal approximation, the test statistic is $z = \frac{14-20}{3.46} = -1.73$. The actual value of α is 0.054 and the rejection region is $X \leq 15$ and the cumulative binomial probability $P(X \leq 15)$ is used. If the normal approximation is used, you would also reject since $z = -1.73$ is in the rejection region which is $Z < -1.645$. Note that if the binomial distribution is used to perform the test, the test statistic has a binomial distribution. If the normal distribution is used to test the hypothesis, the test statistic, Z, has a standard normal distribution.



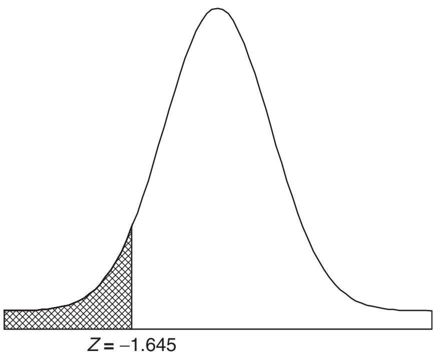


Fig. 10-5 Comparison of the exact test on the left (Binomial) and the approximate test on the right (standard normal).

10.7 The p-value for a test of hypothesis is defined to be the smallest level of significance at which the null hypothesis is rejected. This problem illustrates the computation of the p-value for a statistical test. Use the data in Problem 9.6 to test the null hypothesis that the mean height of all the trees on the farm equals 5 feet (ft) versus the alternative hypothesis that the mean height is less than 5 ft. Find the p-value for this test.

SOLUTION

The computed value for z is $z = (59.22 - 60)/1.01 = -0.77$. The smallest level of significance at which the null hypothesis would be rejected is p-value $= P(z < -0.77) = 0.5 - 0.2794 = 0.2206$. The null hypothesis is rejected if the p-value is less than the pre-set level of significance. In this problem, if the level of significance is pre-set at 0.05, then the null hypothesis is not rejected. The MINITAB solution is as follows where the subcommand `Alternative-1` indicates a lower-tail test.


```
MTB > ZTest mean = 60 sd = 10.111 data in cl ;
SUBC> Alternative -1.
```

Z-Test

```
Test of mu = 60.00 vs mu < 60.00
The assumed sigma = 10.1
```

Variable	N	Mean	StDev	SE Mean	Z	P
height	100	59.22	10.11	1.01	**−0.77**	**0.22**

10.8 A random sample of 33 individuals who listen to talk radio was selected and the hours per week that each listens to talk radio was determined. The data are as follows.

9	8	7	4	8	6	8	8	7	10	8	10	6	7	7	8	9
6	5	8	5	6	8	7	8	5	5	8	7	6	6	4	5	

Test the null hypothesis that $\mu = 5$ hours (h) versus the alternative hypothesis that $\mu \neq 5$ at level of significance $\alpha = 0.05$ in the following three equivalent ways:

(*a*) Compute the value of the test statistic and compare it with the critical value for $\alpha = 0.05$.

(*b*) Compute the p-value corresponding to the computed test statistic and compare the p-value with $\alpha = 0.05$.

(*c*) Compute the $1 - \alpha = 0.95$ confidence interval for μ and determine whether 5 falls in this interval.

SOLUTION

In the following MINITAB output, the standard deviation is found first, and then specified in the `Ztest` statement and the `Zinterval` statement.

```
MTB > standard deviation c1
Standard deviation of hours = 1.6005

MTB > ZTest 5.0 1.6005 'hours';
SUBC>  Alternative 0.
```

Z-Test

```
Test of mu = 5.000 vs mu not = 5.000
The assumed sigma = 1.60

 Variable       N      Mean     StDev    SE Mean        Z        P
 hours         33     6.897     1.600     0.279      6.81   0.0000

MTB > ZInterval 95.0 1.6005 'hours'.
Variable        N      Mean     StDev    SE Mean          95.0 % CI
 hours         33     6.897     1.600     0.279     (  6.351, 7.443)
```

(*a*) The computed value of the test statistic is $Z = \frac{6.897 - 5}{0.279} = 6.81$, the critical values are ± 1.96, and the null hypothesis is rejected. Note that this is the computed value shown in the MINITAB output.

(*b*) The computed *p*-value from the MINITAB output is 0.0000 and since the *p*-value $< \alpha = 0.05$, the null hypothesis is rejected.

(*c*) Since the value specified by the null hypothesis, 5, is not contained in the 95% confidence interval for μ, the null hypothesis is rejected.

These three procedures for testing a null hypothesis against a two-tailed alternative are equivalent.

10.9 The breaking strengths of cables produced by a manufacturer have a mean of 1800 pounds (lb) and a standard deviation of 100 lb. By a new technique in the manufacturing process, it is claimed that the breaking strength can be increased. To test this claim, a sample of 50 cables is tested and it is found that the mean breaking strength is 1850 lb. Can we support the claim at the 0.01 significance level?

SOLUTION

We have to decide between the two hypotheses:

$H_0 : \mu = 1800$ lb, and there is really no change in breaking strength.

$H_1 : \mu > 1800$ lb, and there is a change in breaking strength.

A one-tailed test should be used here; the diagram associated with this test is identical with Fig. 10-4 of Problem 10.5(*a*). At the 0.01 significance level, the decision rule is:

If the z score observed is greater than 2.33, the results are significant at the 0.01 level and H_0 is rejected.

Otherwise, H_0 is accepted (or the decision is withheld).

Under the hypothesis that H_0 is true, we find that

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} = \frac{1850 - 1800}{100/\sqrt{50}} = 3.55$$

which is greater than 2.33. Hence we conclude that the results are *highly significant* and that the claim should thus be supported.

10.10 A group of 50 Internet shoppers were asked how much they spent per year on the Internet. Their responses are shown in Table 10.2. It is desired to test that they spend \$325 per year versus it is different from \$325. Find the *p*-value for the test of hypothesis. What is your conclusion for $\alpha = 0.05$?

Table 10.2

418	379	77	212	378
363	434	348	245	341
331	356	423	330	247
351	151	220	383	257
307	297	448	391	210
158	310	331	348	124
523	356	210	364	406
331	364	352	299	221
466	150	282	221	432
366	195	96	219	202

SOLUTION

The mean of the data is 304.60, the standard deviation is 101.51, the computed test statistic is $z = \dfrac{304.60 - 325}{101.50/\sqrt{50}} = -1.43$. The Z statistic has an approximate standard normal distribution. The computed *p*-value is the following $P(Z < -|\text{computed test statistic}|)$ or $Z > |\text{computed test statistic}|)$ or $P(Z < -1.43) + P(Z > 1.43)$. The answer may be found using Appendix II, or using EXCEL. Using EXCEL, the *p*-value $= 2^*$ NORMSDIST$(-1.43) = 0.1527$, since the normal curve is symmetrical and the areas to the left of -1.43 and to the right of 1.43 are the same, we may simply double the area to the left of -1.43. Because the *p*-value is not less than 0.05, we do not reject the null hypothesis. Refer to Fig. 10-6 to see the graphically the computed *p*-value for this problem.

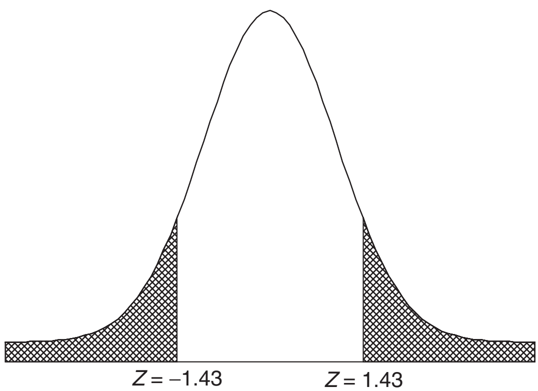


Fig. 10-6 The *p*-value is the sum of the area to the left of $Z = -1.43$ and the area to the right of $Z = 1.43$.

10.11 Refer to Problem 10.10. Use the statistical software MINITAB to analyze the data. Note that the software gives the *p*-value and you, the user, are left to make a decision about the hypothesis based on whatever value you may have assigned to α.

SOLUTION

The pull-down "**Stat ⇒ Basic statistics ⇒ 1 sample Z**" gives the following analysis. The test statistic and the p-value are computed for you.

One-Sample Z: Amount

```
Test of mu = 325 vs not = 325
The assumed standard deviation = 101.51

Variable    N     Mean    StDev   SE Mean       Z      P
Amount     50  304.460  101.508    14.356   -1.43  0.152
```

Note that the software gives the value of the test statistic (−1.43) and the p-value (0.152).

10.12 A survey of individuals who use computer software to file their tax returns is shown in Table 10.3. The recorded response is the time spent in filing their tax returns. The null hypothesis is $H_0 : \mu = 8.5$ hours versus the alternative hypothesis $H_1 : \mu < 8.5$. Find the p-value for the test of hypothesis. What is your conclusion for $\alpha = 0.05$?

Table 10.3

6.2	4.8	8.9	5.6	6.5
11.5	8.6	6.2	8.5	5.2
2.7	14.9	11.2	6.9	7.9
4.8	9.5	12.4	9.7	10.7
8.0	11.8	7.4	9.1	4.9
9.1	6.4	9.5	7.6	6.7
2.6	3.5	6.4	4.3	7.9
3.3	10.3	3.2	11.5	1.7
10.4	8.5	10.8	6.9	5.3
4.9	4.4	9.4	5.6	7.0

SOLUTION

The mean of the data in Table 10.3 is 7.42 h, the standard deviation is 2.91 h and the computed test statistic is $Z = \dfrac{7.42 - 8.5}{2.91/\sqrt{50}} = -2.62$. The Z statistic has an approximate standard normal distribution. The MINITAB pull-down menu "**Calc ⇒ Probability distribution ⇒ Normal**" gives the dialog box shown in Fig. 10-7. The dialog box is filled in as shown.

The output given by the dialog box in Fig. 10-7 is as follows.

Cumulative Distribution Function

```
Normal with mean = 0 and standard deviation = 1

x       P( X <= x )
-2.62  0.0043965
```

The p-value is 0.0044 and since the p-value $< \alpha$, the null hypothesis is rejected. Refer to Fig. 10-8 to see graphically the computed value of the p-value for this problem.

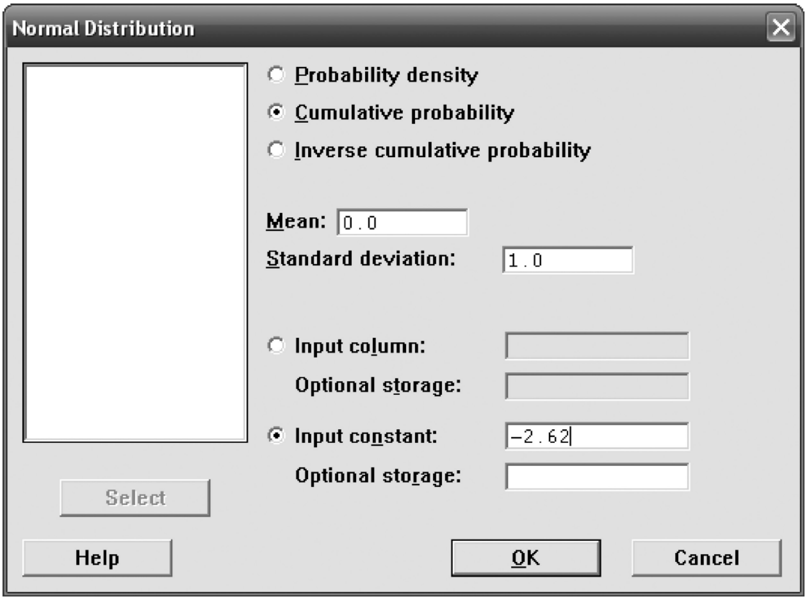


Fig. 10-7 Dialog box for figuring the p-value when the test statistic equals -2.62.

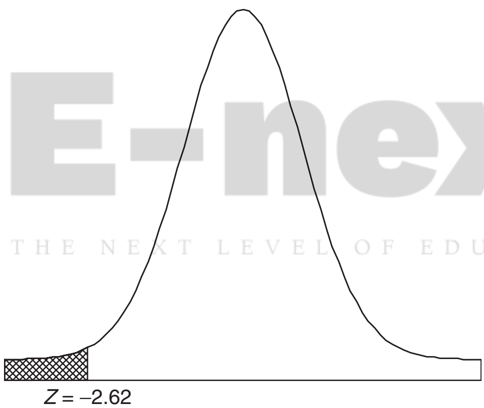


Fig. 10-8 The p-value is the area to the left of $Z = -2.62$.

10.13 Refer to Problem 10.12. Use the statistical software SAS to analyze the data. Note that the software gives the p-value and you, the user, are left to make a decision about the hypothesis based on whatever value you may have assigned to α.

SOLUTION

The SAS output is shown below. The p-value is shown as $\text{Prob} > z = 0.0044$, the same value as was obtained in Problem 10.12. It is the area under the standard normal curve to the left of -2.62. Compare the other quantities in the SAS output with those in Problem 10.12.

SAS OUTPUT:

```
One Sample Z Test for a Mean
Sample Statistics for time
          N          Mean           Std. Dev.    Std. Error
-----------------------------------------------------------
         50          7.42              2.91          0.41
```

```
Hypothesis Test
Null hypothesis:       Mean of time => 8.5
Alternative:           Mean of time <  8.5
with a specified known standard deviation of 2.91
                 Z Statistic       Prob > Z
                 --------------------
                   -2.619          0.0044
  95% Confidence Interval for the Mean
  (Upper Bound Only)
                 Lower Limit       Upper Limit
                 --------------------
                  -infinity        8.10
```

Note that the 95% one-sided interval $(-\infty, 8.10)$ does not contain the null value, 8.5. This is another sign that the null hypothesis should be rejected at the $\alpha = 0.05$ level.

10.14 It is claimed that the average time spent listening to MP3's by those who listen to these devices is 5.5 h per week versus the average time is greater than 5.5. Table 10.4 gives the time spent listening to an MP3 player for 50 individuals. Test $H_0 : \mu = 5.5$ h versus the alternative $H_1 : \mu > 5.5$ h. Find the p-value for the test of hypothesis using the computer software STATISTIX. What is your conclusion for $\alpha = 0.05$?

Table 10.4

6.4	6.4	6.8	7.6	6.9
5.8	5.9	6.9	5.9	6.0
6.3	5.5	6.1	6.4	4.8
6.3	4.2	6.2	5.0	5.9
6.5	6.8	6.8	5.1	6.5
6.7	5.4	5.9	3.5	4.4
6.9	6.7	6.4	5.1	5.4
4.7	7.0	6.0	5.8	5.8
5.7	5.2	4.9	6.6	8.2
6.9	5.5	5.2	3.3	8.3

SOLUTION

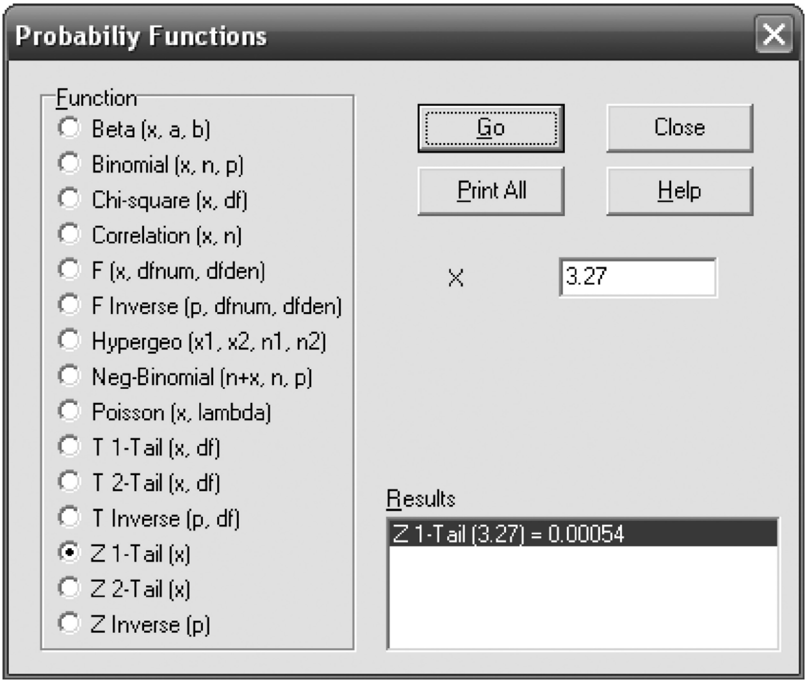
The STATISTIX package gives

```
Statistix 8.0
Descriptive Statistics
Variable       N         Mean              SD
MP3            50        5.9700            1.0158
```

The computed test statistic is $Z = \dfrac{5.97 - 5.5}{1.0158/\sqrt{50}} = 3.27$. The p-value is computed in Fig. 10-9.

Figure 10-10 shows graphically the computed value of the p-value for this problem.

The computed p-value is 0.00054 and since this is less than 0.05 the null hypothesis is rejected.



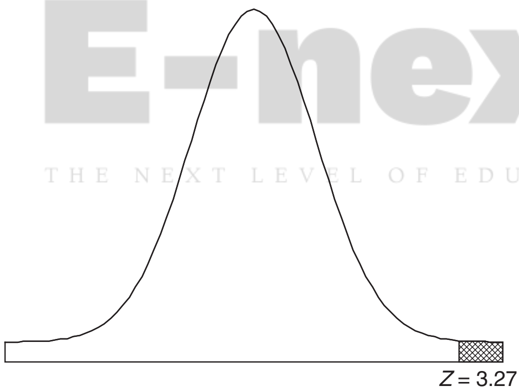


Fig. 10-10 The p-value is the area to the right of $Z = 3.27$

10.15 Using SPSS and the pull-down "**Analyze ⇒ Compare means ⇒ one-sample t test**" and the data in Problem 10.14, test $H_0 : \mu = 5.5$ h versus the alternative $H_0 : \mu > 5.5$ h at $\alpha = 0.05$ by finding the p-value and comparing it with α.

SOLUTION

The SPSS output is as follows:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
MPE	50	5.9700	1.0184	.14366

One-Sample Test

	Test Value=5.5					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Diffrence Lower	Upper
MPE	3.272	49	0.002	.47000	.1813	.7587

In the first portion of the SPSS output, the needed statistics are given. Note that the computed test statistic is referred to as t, not z. This is because for $n > 30$, the t and the z distribution are very similar. The t distribution has a parameter called the degrees of freedom that is equal to $n - 1$. The p-value computed by SPSS is always a two-tailed p-value and is referred to as Sig.(2-tailed). It is equal to 0.002. The 1-tailed value is $0.002/2 = 0.001$. This is close to the value found in Problem 10.14 which equals 0.00054. When using computer software, the user must become aware of the idiosyncrasies of the software.

CONTROL CHARTS

10.16 A control chart is used to control the amount of mustard put into containers. The mean amount of fill is 496 grams (g) and the standard deviation is 5 g. To determine if the machine filling the mustard containers is in proper working order, a sample of 5 is taken every hour for all eight h in the day. The data for two days is given in Table 10.5.

(*a*) Design a decision rule whereby one can be fairly certain that the mean fill is being maintained at 496 g with a standard deviation equal to 5 g for the two days.

(*b*) Show how to graph the decision rule in part (*a*).

Table 10.5

1	2	3	4	5	6	7	8
492.2	486.2	493.6	508.6	503.4	494.9	497.5	490.5
487.9	489.5	503.2	497.8	493.4	492.3	497.0	503.0
493.8	495.9	486.0	493.4	493.9	502.9	493.8	496.4
495.4	494.1	498.4	495.8	493.8	502.8	497.1	489.7
491.7	494.0	496.5	508.0	501.3	498.9	488.3	492.6

9	10	11	12	13	14	15	16
492.2	486.2	493.6	508.6	503.4	494.9	497.5	490.5
487.9	489.5	503.2	497.8	493.4	492.3	497.0	503.0
493.8	495.9	486.0	493.4	493.9	502.9	493.8	496.4
495.4	494.1	498.4	495.8	493.8	502.8	497.1	489.7
491.7	494.0	496.5	508.0	501.3	498.9	488.3	492.6

SOLUTION

(*a*) With 99.73% confidence we can say that the sample mean $\bar{x}$ must lie in the range $\mu_{\bar{x}} - 3\sigma_{\bar{x}}$ to $\mu_{\bar{x}} + 3\sigma_{\bar{x}}$ or $\mu - 3\dfrac{\sigma}{\sqrt{n}}$ to $\mu + 3\dfrac{\sigma}{\sqrt{n}}$. Since $\mu = 496$, $\sigma = 5$, and $n = 5$, it follows that with 99.73% confidence, the sample mean should fall in the interval $496 - 3\dfrac{5}{\sqrt{5}}$ to $496 + 3\dfrac{5}{\sqrt{5}}$ or between 489.29 and 502.71. Hence our decision rule is as follows:

If a sample mean falls inside the range 489.29 g to 502.71 g, assume the machine fill is correct.

Otherwise, conclude that the filling machine is not in proper working order and seek to determine the reasons for the incorrect fills.

(*b*) A record of the sample means can be kept by the use of a chart such as shown in Fig. 10-11, called a *quality control chart*. Each time a sample mean is computed, it is represented by a particular point. As long as the points lie between the lower limit and the upper limit, the process is under control. When a point goes outside these control limits, there is a possibility that something is wrong and investigation is warranted.

The 80 observations are entered into column C1. The pull down menu **"Stat ⇒ Control Charts ⇒ Variable charts for subgroups ⇒ Xbar"** gives a dialog box which when filled in gives the control chart shown in Fig. 10-11.

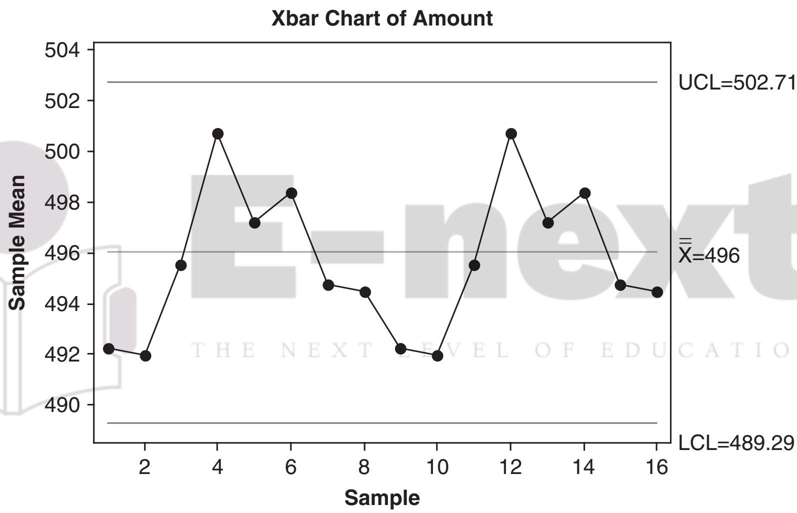


Fig. 10-11 Control chart with 3σ limits for controlling the mean fill of mustard containers.

The control limits specified above are called the 99.73% confidence limits, or briefly, the 3σ limits. Other confidence limits (such as 99% or 95% limits) could be determined as well. The choice in each case depends on the particular circumstances.

TESTS INVOLVING DIFFERENCES OF MEANS AND PROPORTIONS

10.17 An examination was given to two classes consisting of 40 and 50 students, respectively. In the first class the mean grade was 74 with a standard deviation of 8, while in the second class the mean grade was 78 with a standard deviation of 7. Is there a significant difference between the performance of the two classes at the (*a*) 0.05 and (*b*) 0.01 levels?

SOLUTION

Suppose that the two classes come from two populations having the respective means μ_1 and μ_2. We thus need to decide between the hypotheses:

$H_0 : \mu_1 = \mu_2$, and the difference is due merely to chance.

$H_1 : \mu_1 \neq \mu_2$, and there is a significant difference between the classes.

Under hypothesis H_0, both classes come from the same population. The mean and standard deviation of the difference in means are given by

$$\mu_{\bar{X}1-\bar{X}2} = 0 \qquad \text{and} \qquad \sigma_{\bar{X}1-\bar{X}2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} = \sqrt{\frac{8^2}{40} + \frac{7^2}{50}} = 1.606$$

where we have used the sample standard deviations as estimates of σ_1 and σ_2. Thus

$$z = \frac{\bar{X}_1 - \bar{X}}{\sigma_{\bar{X}1-\bar{X}2}} = \frac{74 - 78}{1.606} = -2.49$$

(*a*) For a two-tailed test, the results are significant at the 0.05 level if z lies outside the range -1.96 to 1.96. Hence we conclude that at the 0.05 level there is a significant difference in performance between the two classes and that the second class is probably better.

(*b*) For a two-tailed test, the results are significant at the 0.01 level if z lies outside the range -2.58 and 2.58. Hence we conclude that at the 0.01 level there is no significant difference between the classes.

Since the results are significant at the 0.05 level but not at the 0.01 level, we conclude that the results are *probably significant* (according to the terminology used at the end of Problem 10.5).

10.18 The mean height of 50 male students who showed above-average participation in college athletics was 68.2 inches (in) with a standard deviation of 2.5 in, while 50 male students who showed no interest in such participation had a mean height of 67.5 in with a standard deviation of 2.8 in. Test the hypothesis that male students who participate in college athletics are taller than other male students.

SOLUTION

We must decide between the hypotheses:

$H_0 : \mu_1 = \mu_2$, and there is no difference between the mean heights.

$H_1 : \mu_1 > \mu_2$, and the mean height of the first group is greater than that of the second group.

Under hypothesis H_0,

$$\mu_{\bar{X}1-\bar{X}2} = 0 \qquad \text{and} \qquad \sigma_{\bar{X}1-\bar{X}2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} = \sqrt{\frac{(2.5)^2}{50} + \frac{(2.8)^2}{50}} = 0.53$$

where we have used the sample standard deviations as estimates of σ_1 and σ_2. Thus

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}1-X2}} = \frac{68.2 - 67.5}{0.53} = 1.32$$

Using a one-tailed test at the 0.05 significance level, we would reject hypothesis H_0 if the z score were greater than 1.645. Thus we cannot reject the hypothesis at this level of significance.

It should be noted, however, that the hypothesis can be rejected at the 0.10 level if we are willing to take the risk of being wrong with a probability of 0.10 (i.e., 1 chance in 10).

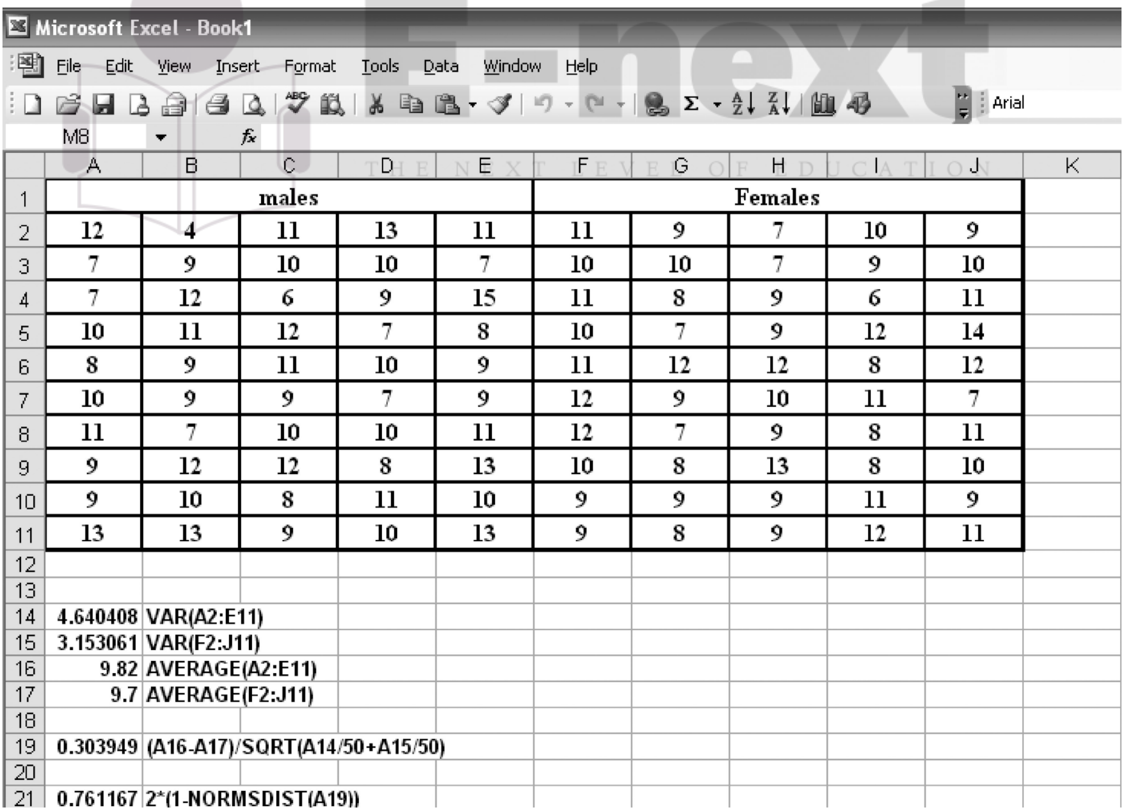
10.19 A study was undertaken to compare the mean time spent on cell phones by male and female college students per week. Fifty male and 50 female students were selected from Midwestern University and the number of hours per week spent talking on their cell phones determined. The results in hours are shown in Table 10.6. It is desired to test $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 \neq 0$ based on these samples. Use EXCEL to find the p-value and reach a decision about the null hypothesis.

Table 10.6 Hours spent talking on cell phone for males and females at Midwestern University

Males					Females				
12	4	11	13	11	11	9	7	10	9
7	9	10	10	7	10	10	7	9	10
7	12	6	9	15	11	8	9	6	11
10	11	12	7	8	10	7	9	12	14
8	9	11	10	9	11	12	12	8	12
10	9	9	7	9	12	9	10	11	7
11	7	10	10	11	12	7	9	8	11
9	12	12	8	13	10	8	13	8	10
9	10	8	11	10	9	9	9	11	9
13	13	9	10	13	9	8	9	12	11

SOLUTION

The data in Table 10.6 are entered into an EXCEL worksheet as shown in Fig. 10-12. The male data are entered into A2 : E11 and the female data are entered into F2:J11. The variance of the male data is computed by entering =VAR(A2 : E11) into A14. The variance of female data is computed by entering =VAR(F2 : J11) into A15. The mean of male data is computed by entering =AVERAGE(A2 : E11)


Microsoft Excel - Book1

	A	B	C	D	E	F	G	H	I	J	K
1	males					Females					
2	12	4	11	13	11	11	9	7	10	9	
3	7	9	10	10	7	10	10	7	9	10	
4	7	12	6	9	15	11	8	9	6	11	
5	10	11	12	7	8	10	7	9	12	14	
6	8	9	11	10	9	11	12	12	8	12	
7	10	9	9	7	9	12	9	10	11	7	
8	11	7	10	10	11	12	7	9	8	11	
9	9	12	12	8	13	10	8	13	8	10	
10	9	10	8	11	10	9	9	9	11	9	
11	13	13	9	10	13	9	8	9	12	11	
12											
13											
14	4.640408	VAR(A2:E11)									
15	3.153061	VAR(F2:J11)									
16	9.82	AVERAGE(A2:E11)									
17	9.7	AVERAGE(F2:J11)									
18											
19	0.303949	(A16-A17)/SQRT(A14/50+A15/50)									
20											
21	0.761167	2*(1-NORMSDIST(A19))									

Fig. 10-12 EXCEL worksheet for computing the p-value in problem 10-19.

into A16. The mean of female data is computed by entering =AVERAGE(F2:J11) into A17. The test statistic is =(A16-A17)/SQRT(A14/50+A15/50) and is shown in A19. The test statistic has a standard normal distribution and a value of 0.304. The expression =2*(1-NORMSDIST(A19)) computes the area to the right of 0.304 and doubles it. This gives a p-value of 0.761.

Since the p-value is not smaller than any of the usual α values such as 0.01 or 0.05, the null hypothesis is not rejected. The probability of obtaining samples like the one we obtained is 0.761, assuming the null hypothesis to be true. Therefore, there is no evidence to suggest that the null hypothesis is false and that it should be rejected.

10.20 Two groups, A and B, consist of 100 people each who have a disease. A serum is given to group A but not to group B (which is called the *control*); otherwise, the two groups are treated identically. It is found that in groups A and B, 75 and 65 people, respectively, recover from the disease. At significance levels of (*a*) 0.01, (*b*) 0.05, and (*c*) 0.10, test the hypothesis that the serum helps cure the disease. Compute the p-value and show that p-value>0.01, p-value>0.05, but p-value<0.10.

SOLUTION

Let p_1 and p_2 denote the population proportions cured by (1) using the serum and (2) not using the serum, respectively. We must decide between two hypotheses:

$H_0: p_1 = p_2$, and the observed differences are due to chance (i.e. the serum is ineffective).

$H_1: p_1 > p_2$, and the serum is effective.

Under hypothesis H_0,

$$\mu_{P1-P2} = 0 \qquad \text{and} \qquad \sigma_{P1-P2} = \sqrt{pq\left(\frac{1}{N_1}+\frac{1}{N_2}\right)} = \sqrt{(0.70)(0.30)\left(\frac{1}{100}+\frac{1}{100}\right)} = 0.0648$$

where we have used as an estimate of p the average proportion of cures in the two sample groups given by $(75+65)/200 = 0.70$, and where $q = 1 - p = 0.30$. Thus

$$z = \frac{P_1 - P_2}{\sigma_{P1-P2}} = \frac{0.750 - 0.650}{0.0648} = 1.54$$

(*a*) Using a one-tailed test at the 0.01 significance level, we would reject hypothesis H_0 only if the z score were greater than 2.33. Since the z score is only 1.54, we must conclude that the results are due to chance at this level of significance.

(*b*) Using a one-tailed test at the 0.05 significance level, we would reject H_0 only if the z score were greater than 1.645. Hence we must conclude that the results are due to chance at this level also.

(*c*) If a one-tailed test at the 0.10 significance level were used, we would reject H_0 only if the z score were greater than 1.28. Since this condition is satisfied, we conclude that the serum is effective at the 0.10 level.

(*d*) Using EXCEL, the p-value is given by =1 - NORMSDIST(1.54) which equals 0.06178. This is the area to the right of 1.54. Note that the p-value is greater than 0.01, 0.05, but is less than 0.10.

Note that these conclusions depend how much we are willing to risk being wrong. If the results are actually due to chance, but we conclude that they are due to the serum (Type I error), we might proceed to give the serum to large groups of people—only to find that it is actually ineffective. This is a risk that we are not always willing to assume.

On the other hand, we could conclude that the serum does not help, whereas it actually does help (Type II error). Such a conclusion is very dangerous, especially if human lives are at stake.

10.21 Work Problem 10.20 if each group consists of 300 people and if 225 people in group A and 195 people in group B are cured. Find the p-value using EXCEL and comment on your decision.

SOLUTION

Note that in this case the proportions of people cured in the two groups are $225/300 = 0.750$ and $195/300 = 0.650$, respectively, which are the same as in Problem 10.20. Under hypothesis H_0,

$$\mu_{P1-P2} = 0 \qquad \text{and} \qquad \sigma_{P1-P2} = \sqrt{pq\left(\frac{1}{N_1} + \frac{1}{N_2}\right)} = \sqrt{(0.70)(0.30)\left(\frac{1}{300} + \frac{1}{300}\right)} = 0.0374$$

where $(225 + 195)/600 = 0.70$ is used as an estimate of p. Thus

$$z = \frac{P_1 - P_2}{\sigma_{P1-P2}} = \frac{0.750 - 0.650}{0.0374} = 2.67$$

Since this value of z is greater than 2.33, we can reject the hypothesis at the 0.01 significance level; that is, we can conclude that the serum is effective with only a 0.01 probability of being wrong.

This shows how increasing the sample size can increase the reliability of decisions. In many cases, however, it may be impractical to increase sample sizes. In such cases we are forced to make decisions on the basis of available information and must therefore contend with greater risks of incorrect decisions.

p-value = 1 - NORMSDIST(2.67) = 0.003793. This is less than 0.01.

10.22 A sample poll of 300 voters from district A and 200 voters from district B showed that 56% and 48%, respectively, were in favor of a given candidate. At a significance level of 0.05, test the hypotheses (a) that there is a difference between the districts and (b) that the candidate is preferred in district A (c) calculate the p-value for parts (a) and (b).

SOLUTION

Let p_1 and p_2 denote the proportions of all voters from districts A and B, respectively, who are in favor of the candidate. Under the hypothesis $H_0 : p_1 = p_2$, we have

$$\mu_{P1-P2} = 0 \qquad \text{and} \qquad \sigma_{P1-P2} = \sqrt{pq\left(\frac{1}{N_1} + \frac{1}{N_2}\right)} = \sqrt{(0.528)(0.472)\left(\frac{1}{300} + \frac{1}{200}\right)} = 0.0456$$

where we have used as estimates of p and q the values $[(0.56)(300) + (0.48)(200)]/500 = 0.528$ and $(1 - 0.528) = 0.472$, respectively. Thus

$$z = \frac{P_1 - P_2}{\sigma_{P1-P2}} = \frac{0.560 - 0.480}{0.0456} = 1.75$$

(a) If we wish only to determine whether there is a difference between the districts, we must decide between the hypotheses $H_0 : p_1 = p_2$ and $H_1 : p_1 \neq p_2$, which involves a two-tailed test. Using a two-tailed test at the 0.05 significance level, we would reject H_0 if z were outside the interval -1.96 to 1.96. Since $z = 1.75$ lies inside this interval, we cannot reject H_0 at this level; that is, there is no significant difference between the districts.

(b) If we wish to determine whether the candidate is preferred in district A, we must decide between the hypotheses $H_0 : p_1 = p_2$ and $H_1 : p_1 > p_2$, which involves a one-tailed test. Using a one-tailed test at the 0.05 significance level, we would reject H_0 if z were greater than 1.645. Since this is the case, we can reject H_0 at this level and conclude that the candidate is preferred in district A.

(c) For the two-tailed alternative, the p-value = 2*(1-NORMSDIST(1.75)) = 0.0801. You cannot reject the null hypothesis at $\alpha = 0.05$. For the one-tailed alternative, p-value = 1-NORMSDIST(1.75) = 0.04006. You can reject the null hypothesis at alpha = 0.05.

TESTS INVOLVING BINOMIAL DISTRIBUTIONS

10.23 An instructor gives a short quiz involving 10 true-false questions. To test the hypothesis that students are guessing, the instructor adopts the following decision rule:

If seven or more answers are correct, the student is not guessing.
If less than seven answers are correct, the student is guessing.

Find the following probability of rejecting the hypothesis when it is correct using (*a*) the binomial probability formula and (*b*) EXCEL.

SOLUTION

(*a*) Let p be the probability that a question is answered correctly. The probability of getting X problems out of 10 correct is $\binom{10}{X}p^X q^{10-X}$, where $q = 1 - p$. Then under the hypothesis $p = 0.5$ (i.e., the student is guessing),

$$\Pr\{7 \text{ or more correct}\} = \Pr\{7 \text{ correct}\} + \Pr\{8 \text{ correct}\} + \Pr\{9 \text{ correct}\} + \Pr\{10 \text{ correct}\}$$

$$= \binom{10}{7}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3 + \binom{10}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 + \binom{10}{9}\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right) + \binom{10}{10}\left(\frac{1}{2}\right)^{10} = 0.1719$$

Thus the probability of concluding that students are not guessing when in fact they are guessing is 0.1719. Note that this is the probability of a Type I error.

(*b*) Enter the numbers 7, 8, 9, and 10 into A1:A4 of the EXCEL worksheet. Next enter = BINOMDIST(A1, 10, 0.5, 0). Next perform a click-and-drag from B1 to B4. In B5 enter = SUM(B1 : B4). The answer appears in B5.

A	B
7	0.117188
8	0.043945
9	0.009766
10	0.000977
	0.171875

10.24 In Problem 10.23, find the probability of accepting the hypothesis $p = 0.5$ when actually $p = 0.7$. Find the answer (*a*) using the binomial probability formula and (*b*) using EXCEL.

SOLUTION

(*a*) Under the hypothesis $p = 0.7$,

$$\Pr\{\text{less than 7 correct}\} = 1 - \Pr\{7 \text{ or more correct}\}$$

$$= 1 - \left[\binom{10}{7}(0.7)^7(0.3)^3 + \binom{10}{8}(0.7)^8(0.3)^2 + \binom{10}{9}(0.7)^9(0.3) + \binom{10}{10}(0.3)^{10}\right]$$

$$= 0.3504$$

(*b*) The EXCEL solution is:

Pr{less than 7 correct when $p = 0.7$} is given by = BINOMDIST(6,10,0.7,1) which equals 0.350389. The 1 in the function BINOMDIST says to accumulate from 0 to 6 the binomial probabilities with $n = 10$ and $p = 0.7$.

10.25 In Problem 10.23, find the probability of accepting the hypothesis $p = 0.5$ when actually (*a*) $p = 0.6$, (*b*) $p = 0.8$, (*c*) $p = 0.9$, (*d*) $p = 0.4$, (*e*) $p = 0.3$, (*f*) $p = 0.2$, and (*g*) $p = 0.1$.

SOLUTION

(*a*) If $p = 0.6$,

$$\text{Required probability} = 1 - [\Pr\{7 \text{ correct}\} + \Pr\{8 \text{ correct}\} + \Pr\{9 \text{ correct}\} + \Pr\{10 \text{ correct}\}]$$

$$= 1 - \left[\binom{10}{7}(0.6)^7(0.4)^3 + \binom{10}{8}(0.6)^8(0.4)^2 + \binom{10}{9}(0.6)^9(0.4) + \binom{10}{10}(0.6)^{10}\right] = 0.618$$

The results for parts (*b*) through (*g*) can be found similarly and are shown in Table 10.7, together with the values corresponding to $p = 0.5$ and to $p = 0.7$. Note that the probability is denoted in Table 10.7 by β (probability of a Type II error); the β entry for $p = 0.5$ is given by $\beta = 1 - 0.1719 = 0.828$ (from Problem 10.23), and the β entry for $p = 0.7$ is from Problem 10.24.

Table 10.7

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
β	1.000	0.999	0.989	0.945	0.828	0.618	0.350	0.121	0.013

10.26 Use Problem 10.25 to construct the graph of β versus *p*.

SOLUTION

The required graph is shown in Fig. 10-13.

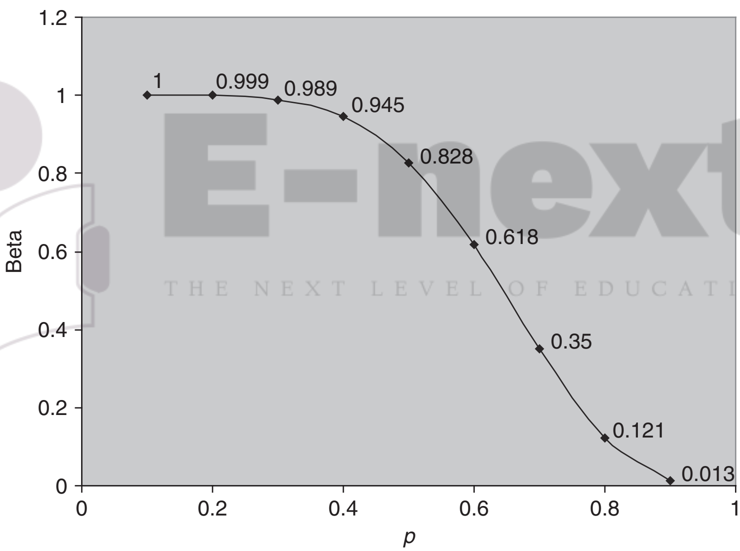


Fig. 10-13 Graph of Type II errors for Problem 10.25.

10.27 The null hypothesis is that a die is fair and the alternative hypothesis is that the die is biased such that the face six is more likely to occur than it would be if the die were fair. The hypothesis is tested by rolling it 18 times and observing how many times the face six occurs. Find the *p*-value if the face six occurs 7 times in the 18 rolls of the die.

SOLUTION

Zero through 18 are entered into A1:A19 in the EXCEL worksheet. =BINOMDIST(A1, 18, 0.16666, 0) is entered in B1 and a click-and-drag is performed from B1 to B19 to give the individual binomial probabilities. =BINOMDIST(A1, 18, 0.16666, 1) is entered into C1 and a click-and-drag is performed from C1 to C19 to give the cumulative binomial probabilities.

A	B	C
0	0.037566	0.037566446
1	0.135233	0.17279916
2	0.229885	0.402683738
3	0.245198	0.647882186
4	0.18389	0.831772194
5	0.102973	0.934745656
6	0.04462	0.979365347
7	0.015297	0.994662793
8	0.004207	0.998869389
9	0.000935	0.999804143
10	0.000168	0.999972391
11	2.45E-05	0.999996862
12	2.85E-06	0.999999717
13	2.64E-07	0.99999998
14	1.88E-08	0.999999999
15	1E-09	1
16	3.76E-11	1
17	8.86E-13	1
18	9.84E-15	1

The p-value is $p\{x \geq 7\} = 1 - P\{X \leq 6\} = 1 - 0.979 = 0.021$.The outcome $X = 6$ is significant at $\alpha = 0.05$ but not at $\alpha = 0.01$.

10.28 To test that 40% of taxpayers use computer software when figuring their taxes against the alternative that the percent is greater than 40%, 300 taxpayers are randomly selected and asked if they use computer software. If 131 of 300 use computer software, find the p-value for this observation.

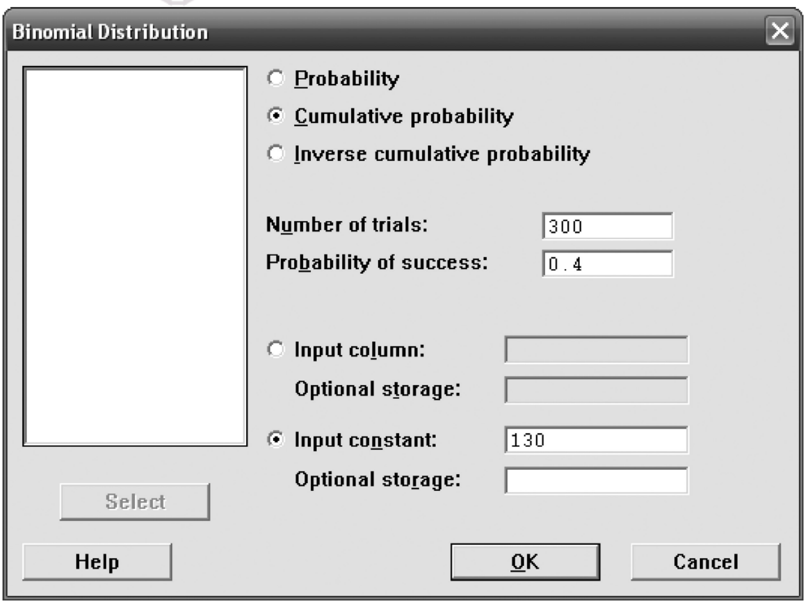


Fig. 10-14 Binomial dialog box for computing 130 or fewer computer software users in 300 when 40% of all taxpayers use computer software.

SOLUTION

The null hypothesis is $H_0: p=0.4$ versus the alternative $H_a: p>0.4$. The observed valued of X is 131, where X is the number who use computer software. The p-value $= P\{X \geq 131$ when $p=0.4\}$. The $p-\text{value} = 1 - P\{X \leq 130 \text{ when } p = 0.4\}$ Using MINITAB, the pull down "**Calc ⇒ Probability Distribution ⇒ Binomial**" gives the dialog box shown in Fig 10-14.

The dialog box in Fig. 10-14 produces the following output.

Cumulative Distribution Function

```
Binomial with n=300 and p=0.4

x        P(X<=x)
130      0.891693
```

The p-value is $1 - P\{X \leq 130 \text{ when } p = 0.4\} = 1 - 0.8971 = 0.1083$. The outcome $X=131$ is not significant at 0.01, 0.05, or 0.10.

Supplementary Problems

TESTS OF MEANS AND PROPORTIONS, USING NORMAL DISTRIBUTIONS

10.29 An urn contains marbles that are either red or blue. To test the null hypothesis of equal proportions of these colors, we agree to sample 64 marbles with replacement, noting the colors drawn, and to adopt the following decision rule:

Accept the null hypothesis if $28 \leq X \leq 36$, where $X=$ number of red marbles in the 64.

Reject the null hypothesis if $X \leq 27$ or if $X \geq 37$.

(*a*) Find the probability of rejecting the null hypothesis when it is correct.

(*b*) Graph the decision rule and the result obtained in part (*a*).

10.30 (*a*) What decision rule would you adopt in Problem 10.29 if you require that the probability of rejecting the hypothesis when it is actually correct be no more than 0.01 (i.e., you want a 0.01 significance level)?

(*b*) At what level of confidence would you accept the hypothesis?

(*c*) What would the decision rule be if the 0.05 significance level were adopted?

10.31 Suppose that in Problem 10.29 we wish to test the hypothesis that there is a *greater proportion* of red than blue marbles.

(*a*) What would you take as the null hypothesis, and what would be the alternative hypothesis?

(*b*) Should you use a one- or a two-tailed test? Why?

(*c*) What decision rule should you adopt if the significance level is 0.05?

(*d*) What is the decision rule if the significance level is 0.01?

10.32 A pair of dice is tossed 100 times and it is observed that 7's appear 23 times. Test the hypothesis that the dice are fair (i.e., not loaded) at the 0.05 significance level by using (*a*) a two-tailed test and (*b*) a one-tailed test. Discuss your reasons, if any, for preferring one of these tests over the other.

10.33 Work Problem 10.32 if the significance level is 0.01.

10.34 A manufacturer claimed that at least 95% of the equipment that she supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test her claim at significance levels of (*a*) 0.01 and (*b*) 0.05.

10.35 The claim is made that Internet shoppers spend on the average \$335 per year. It is desired to test that this figure is not correct at $\alpha = 0.075$. Three hundred Internet shoppers are surveyed and it is found that the sample mean = \$354 and the standard deviation = \$125. Find the value of the test statistic, the critical values, and give your conclusion.

10.36 It has been found from experience that the mean breaking strength of a particular brand of thread is 9.72 ounces (oz) with a standard deviation of 1.40 oz. A recent sample of 36 pieces of this thread showed a mean breaking strength of 8.93 oz. Test the null hypothesis $H_0: \mu = 9.72$ versus the alternative $H_0: \mu < 9.72$ by giving the value of the test statistic and the critical value for $\alpha = 0.10$ and $\alpha = 0.025$. Is the result significant at $\alpha = 0.10$. Is the result significant at $\alpha = 0.025$?

10.37 A study was designed to test the null hypothesis that the average number of e-mails sent weekly by employees in a large city equals 25.5 versus the mean is greater than 25.5. Two hundred employees were surveyed across the city and it was found that $\bar{x} = 30.3$ and $s = 10.5$. Give the value of the test statistic, the critical value for $\alpha = 0.03$, and your conclusion.

10.38 For large n ($n > 30$) and known standard deviation the standard normal distribution is used to perform a test concerning the mean of the population from which the sample was selected. The alternative hypothesis $H_a : \mu < \mu_0$ is called a *lower-tailed alternative* and the alternative hypothesis $H_a : \mu > \mu_0$ is called a *upper-tailed alternative*. For an upper-tailed alternative, give the EXCEL expression for the critical value if $\alpha = 0.1$, $\alpha = 0.01$, and $\alpha = 0.001$.

p-VALUES FOR HYPOTHESES TESTS

10.39 To test a coin for its balance, it is flipped 15 times. The number of heads obtained is 12. Give the *p*-value corresponding to this outcome. Use BINOMDIST of EXCEL to find the *p*-value.

10.40 Give the *p*-value for the outcome in Problem 10.35.

10.41 Give the *p*-value for the outcome in Problem 10.36.

10.42 Give the *p*-value for the outcome in Problem 10.37.

QUALITY CONTROL CHARTS

10.43 In the past a certain type of thread produced by a manufacturer has had a mean breaking strength of 8.64 oz and a standard deviation of 1.28 oz. To determine whether the product is conforming to standards, a sample of 16 pieces of thread is taken every 3 hours and the mean breaking strength is determined. Record the (*a*) 99.73% (or 3σ), (*b*) 99%, and (*c*) 95% control limits on a quality control chart and explain their applications.

10.44 On average, about 3% of the bolts produced by a company are defective. To maintain this quality of performance, a sample of 200 bolts produced is examined every 4 hours. Determine the (*a*) 99% and (*b*) 95% control limits for the number of defective bolts in each sample. Note that only *upper control limits* are needed in this case.

TESTS INVOLVING DIFFERENCES OF MEANS AND PROPORTIONS

10.45 A study compared the mean lifetimes in hours of two types of types of light bulbs. The results of the study are shown in Table 10.8.

Table 10.8

	Environmental Bulb	Traditional Bulb
n	75	75
Mean	1250	1305
Std. dev	55	65

Test $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 \neq 0$ at $\alpha = 0.05$. Give the value of the test statistic and compute the p-value and compare the p-value with $\alpha = 0.05$. Give your conclusion.

10.46 A study compared the grade point averages (GPAS) of 50 high school seniors with a TV in their bedroom with the GPAS of 50 high school seniors without a TV in their bedrooms. The results are shown in Table 10.9. The alternative is that the mean GPA is greater for the group with no TV in their bedroom. Give the value of the test statistic assuming no difference in mean GPAs. Give the p-value and your conclusion for $\alpha = 0.05$ and for $\alpha = 0.10$.

Table 10.9

	TV in Bedroom	No TV in Bedroom
n	50	50
Mean	2.58	2.77
Std. dev	0.55	0.65

10.47 On an elementary school spelling examination, the mean grade of 32 boys was 72 with a standard deviation of 8, while the mean grade of 36 girls was 75 with a standard deviation of 6. The alternative is that the girls are better at spelling than the boys. Give the value of the test statistic assuming no difference in boys and girls at spelling. Give the p-value and your conclusion for $\alpha = 0.05$ and for $\alpha = 0.10$.

10.48 To test the effects of a new fertilizer on wheat production, a tract of land was divided into 60 squares of equal areas, all portions having identical qualities in terms of soil, exposure to sunlight, etc. The new fertilizer was applied to 30 squares and the old fertilizer was applied to the remaining squares. The mean number of bushels (bu) of wheat harvested per square of land using the new fertilizer was 18.2 bu with a standard deviation of 0.63 bu. The corresponding mean and standard deviation for the squares using the old fertilizer were 17.8 and 0.54 bu, respectively. Using significance levels of (*a*) 0.05 and (*b*) 0.01, test the hypothesis that the new fertilizer is better than the old one.

10.49 Random samples of 200 bolts manufactured by machine A and of 100 bolts manufactured by machine B showed 19 and 5 defective bolts, respectively.

(*a*) Give the test statistic, the p-value, and your conclusion at $\alpha = 0.05$ for testing that the two machines show different qualities of performance

(*b*) Give the test statistic, the p-value, and your conclusion at $\alpha = 0.05$ for testing that machine B is performing better than machine A.

10.50 Two urns, A and B, contain equal numbers of marbles, but the proportion of red and white marbles in each of the urns is unknown. A sample of 50 marbles from each urn is selected from each urn with replacement. There are 32 red in the 50 from urn A and 23 red in the 50 from urn B.

(*a*) Test, at $\alpha = 0.05$, that the proportion of red is the same versus the proportion is different by giving the computed test statistic, the computed *p*-value and your conclusion.

(b) Test, at $\alpha = 0.05$, that A has a greater proportion than B by giving the computed test statistic, the *p*-value, and your conclusion.

10.51 A coin is tossed 15 times in an attempt to determine whether it is biased so that heads are more likely to occur than tails. Let $X =$ the number of heads to occur in 15 tosses. The coin is declared biased in favor of heads if $X \geq 11$. Use EXCEL to find α.

10.52 A coin is tossed 20 times to determine if it is unfair. It is declared unfair if $X = 0, 1, 2, 18, 19, 20$ where $X =$ the number of tails to occur. Use EXCEL to find α.

10.53 A coin is tossed 15 times in an attempt to determine whether it is biased so that heads are more likely to occur than tails. Let $X =$ the number of heads to occur in 15 tosses. The coin is declared biased in favor of heads if $X \geq 11$. Use EXCEL find β if $p = 0.6$.

10.54 A coin is tossed 20 times to determine if it is unfair. It is declared unfair if $X = 0, 1, 2, 18, 19, 20$ where $X =$ the number of tails to occur. Use EXCEL to find β if $p = 0.9$.

10.55 A coin is tossed 15 times in an attempt to determine whether it is biased so that heads are more likely to occur than tails. Let $X =$ the number of heads to occur in 15 tosses. The coin is declared biased in favor of heads if $X \geq 11$. Find the *p*-value for the outcome $X = 10$. Compare the *p*-value with the value of α in this problem.

10.56 A coin is tossed 20 times to determine if it is unfair. It is declared unfair if $X = 0, 1, 2, 3, 4, 16, 17, 18, 19, 20$ where $X =$ the number of tails to occur. Find the *p*-value for the outcome $X = 17$. Compare the *p*-value with the value of α in this problem.

10.57 A production line manufactures cell phones. Three percent defective is considered acceptable. A sample of size 50 is selected from a day's production. If more than 3 are found defective in the sample, it is concluded that the defective percent exceeds the 3% figure and the line is stopped until it meets the 3% figure. Use EXCEL to determine α?

10.58 In Problem 10.57, find the probability that a 4% defective line will not be shut down.

10.59 To determine if it is balanced, a die is rolled 20 times. It is declared to be unbalanced so that the face six occurs more often than 1/6 of the time if more than 5 sixes occur in the 20 rolls. Find the value for α. If the die is rolled 20 times and a six occurs 6 times, find the *p*-value for this outcome.