26

Testing hypotheses: elaboration

In the previous chapter we introduced the setup for testing a null hypothesis against an alternative hypothesis using a test statistic T. The notions of type I error and type II error were introduced. A type I error occurs when we falsely reject H_0 on the basis of the observed value of T, whereas a type II error occurs when we falsely do not reject H_0. The decision to reject H_0 or not was based on the size of the p-value. In this chapter we continue the introduction of basic concepts of testing hypotheses, such as *significance level* and *critical region*, and investigate the probability of committing a type II error.

26.1 Significance level

As mentioned in the previous chapter, there is no general rule that specifies a level below which the p-value is considered exceptionally small. However, there are situations where this level is set *a priori*, and the question is: which values of the test statistic should then lead to rejection of H_0? To illustrate this, consider the following example. The speed limit on freeways in The Netherlands is 120 kilometers per hour. A device next to freeway A2 between Amsterdam and Utrecht measures the speed of passing vehicles. Suppose that the device is designed in such a way that it conducts three measurements of the speed of a passing vehicle, modeled by a random sample X_1, X_2, X_3. On the basis of the value of the average $\bar{X}_3$, the driver is either fined for speeding or not. For what values of $\bar{X}_3$ should we fine the driver, if we allow that 5% of the drivers are fined unjustly?

Let us rephrase things in terms of a testing problem. Each measurement can be thought of as

$$\text{measurement} = \text{true speed} + \text{measurement error}.$$

Suppose for the moment that the measuring device is carefully calibrated, so that the measurement error is modeled by a random variable with mean zero and known variance σ^2, say $\sigma^2 = 4$. Moreover, in physical experiments such as this one, the measurement error is often modeled by a random variable with a normal distribution. In that case, the measurements X_1, X_2, X_3 are modeled by a random sample from an $N(\mu, 4)$ distribution, where the parameter μ represents the true speed of the passing vehicle. Our testing problem can now be formulated as testing

$$H_0 : \mu = 120 \quad \text{against} \quad H_1 : \mu > 120,$$

with test statistic

$$T = \frac{X_1 + X_2 + X_3}{3} = \bar{X}_3.$$

Since sums of independent normal random variables again have a normal distribution (see Remark 11.2), it follows that $\bar{X}_3$ has an $N(\mu, 4/3)$ distribution. In particular, the distribution of $T = \bar{X}_3$ is centered around μ no matter what the value of μ is. Values of T close to 120 are therefore in favor of H_0. Values of T that are far from 120 are considered as strong evidence against H_0. Values much larger than 120 suggest that $\mu > 120$ and are therefore in favor of H_1. Values much smaller than 120 suggest that $\mu < 120$. They also constitute evidence against H_0, but even stronger evidence against H_1. Thus we reject H_0 in favor of H_1 *only* for values of T larger than 120. See also Figure 26.1.

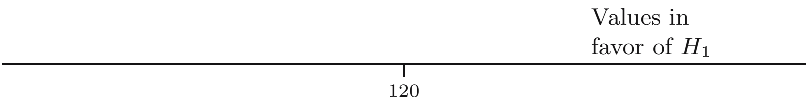


Fig. 26.1. Possible values of $T = \bar{X}_3$.

Rejection of H_0 in favor of H_1 corresponds to fining the driver for speeding. Unjustly fining a driver corresponds to falsely rejecting H_0, i.e., committing a type I error. Since we allow 5% of the drivers to be fined unjustly, we are dealing with a testing problem where the probability of committing a type I error is set a priori at 0.05. The question is: for which values of T should we reject H_0? The decision rule for rejecting H_0 should be such that the corresponding probability of committing a type I error is 0.05. The value 0.05 is called the significance level.

Significance level. The *significance level* is the largest acceptable probability of committing a type I error and is denoted by α, where $0 < \alpha < 1$.

We speak of "performing the test at level α," as well as "rejecting H_0 in favor of H_1 at level α." In our example we are testing $H_0 : \mu = 120$ against $H_1 : \mu > 120$ at level 0.05.

Quick exercise 26.1 Suppose that in the freeway example $H_0 : \mu = 120$ is rejected in favor of $H_1 : \mu > 120$ at level $\alpha = 0.05$. Will it necessarily be rejected at level $\alpha = 0.01$? On the other hand, suppose that $H_0 : \mu = 120$ is rejected in favor of $H_1 : \mu > 120$ at level $\alpha = 0.01$. Will it necessarily be rejected at level $\alpha = 0.05$?

Let us continue with our example and determine for which values of $T = \bar{X}_3$ we should reject H_0 at level $\alpha = 0.05$ in favor of $H_1 : \mu > 120$. Suppose we decide to fine each driver whose recorded average speed is 121 or more, i.e., we reject H_0 whenever $T \geq 121$. Then how large is the probability of a type I error $\mathrm{P}(T \geq 121)$? When $H_0 : \mu = 120$ is true, then $T = \bar{X}_3$ has an $N(120, 4/3)$ distribution, so that by the change-of-units rule for the normal distribution (see page 106), the random variable

$$Z = \frac{T - 120}{2/\sqrt{3}}$$

has an $N(0, 1)$ distribution. This implies that

$$\mathrm{P}(T \geq 121) = \mathrm{P}\left(\frac{T - 120}{2/\sqrt{3}} \geq \frac{121 - 120}{2/\sqrt{3}}\right) = \mathrm{P}(Z \geq 0.87)\,.$$

From Table B.1, we find $\mathrm{P}(Z \geq 0.87) = 0.1922$, which means that the probability of a type I error is greater than the significance level $\alpha = 0.05$. Since this level was defined as the largest acceptable probability of a type I error, we do not reject H_0. Similarly, if we decide to reject H_0 whenever we record an average of 122 or more, the probability of a type I error equals 0.0416 (check this). This is smaller than $\alpha = 0.05$, so in that case we reject H_0. The boundary case is the value c that satisfies $\mathrm{P}(T \geq c) = 0.05$. To find c, we must solve

$$\mathrm{P}\left(Z \geq \frac{c - 120}{2/\sqrt{3}}\right) = 0.05.$$

From Table B.2 we have that $z_{0.05} = t_{\infty,0.05} = 1.645$, so that we find

$$\frac{c - 120}{2/\sqrt{3}} = 1.645,$$

which leads to

$$c = 120 + 1.645 \cdot \frac{2}{\sqrt{3}} = 121.9.$$

Hence, if we set the significance level α at 0.05, we should reject $H_0 : \mu = 120$ in favor of $H_1 : \mu > 120$ whenever $T \geq 121.9$. For our freeway example this means that if the average recorded speed of a passing vehicle is greater than or equal to 121.9, then the driver is fined for speeding. With this decision rule, at most 5% of the drivers get fined unjustly.

In connection with p-values: the significance level is the level below which the p-value is sufficiently small to reject H_0. Indeed, for any observed value $t \geq 121.9$ we reject H_0, and the p-value for such a t is at most 0.05:

$$P(T \geq t) \leq P(T \geq 121.9) = 0.05.$$

We will see more about this relation in the next section.

26.2 Critical region and critical values

In the freeway example the significance level 0.05 corresponds to the decision rule "reject $H_0 : \mu = 120$ in favor $H_1 : \mu > 120$ whenever $T \geq 121.9$." The set $K = [121.9, \infty)$ consisting of values of the test statistic T for which we reject H_0 is called critical region. The value 121.9, which is the boundary case between rejecting and not rejecting H_0, is called the critical value.

> CRITICAL REGION AND CRITICAL VALUES. Suppose we test H_0 against H_1 at significance level α by means of a test statistic T. The set $K \subset \mathbb{R}$ that corresponds to all values of T for which we reject H_0 in favor of H_1 is called the *critical region*. Values on the boundary of the critical region are called *critical values*.

The precise shape of the critical region depends on both the chosen significance level α and the test statistic T that is used. But it will always be such that the probability that $T \in K$ satisfies

$$P(T \in K) \leq \alpha \quad \text{in the case that } H_0 \text{ is true.}$$

At this point it becomes important to emphasize whether probabilities are computed under the assumption that H_0 is true. With a slight abuse of notation, we briefly write $P(T \in K \mid H_0)$ for the probability.

Relation with p-values

If we record average speed $t = 124$, then this value falls in the critical region $K = [121.9, \infty)$, so that $H_0 : \mu = 120$ is rejected in favor $H_1 : \mu > 120$. On the other hand we can also compute the p-value corresponding to the observed value 124. Since values of T to the right provide stronger evidence against H_0, the p-value is the following right tail probability

$$P(T \geq 124 \mid H_0) = P\left(\frac{T - 120}{2/\sqrt{3}} \geq \frac{124 - 120}{2/\sqrt{3}}\right) = P(Z \geq 3.46) = 0.0003,$$

which is smaller than the significance level 0.05. This is no coincidence.

In general, suppose that we perform a test at level α using test statistic T and that we have observed t as the value of our test statistic. Then

$$t \in K \quad \Leftrightarrow \quad \text{the } p\text{-value corresponding to } t \text{ is less than or equal to } \alpha.$$

Figure 26.2 illustrates this for a testing problem where values of T to the right provide evidence against H_0 and in favor of H_1. In that case, the p-value corresponds to the right tail probability $\mathrm{P}(T \geq t \mid H_0)$. The shaded area to the right of c_α corresponds to $\alpha = \mathrm{P}(T \geq c_\alpha \mid H_0)$, whereas the more intensely shaded area to the right of t represents the p-value. We see that deciding whether to reject H_0 at a given significance level α can be done by comparing either t with c_α or the p-value with α. For this reason the p-value is sometimes called the *observed significance level.*

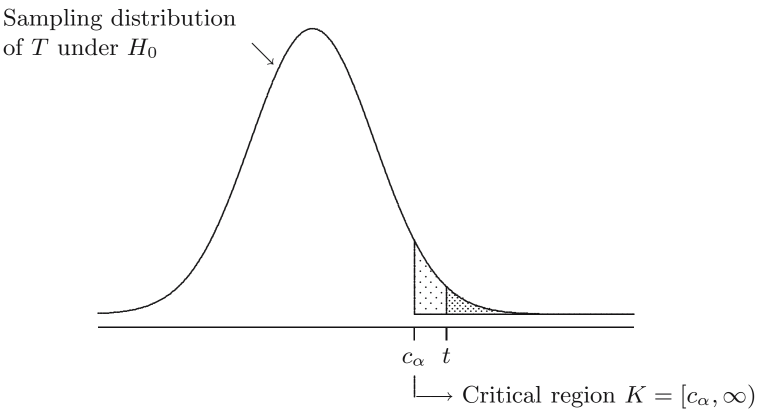


Fig. 26.2. P-value and critical value.

The concepts of critical value and p-value have their own merit. The critical region and the corresponding critical values specify exactly what values of T lead to rejection of H_0 at a given level α. This can be done even without obtaining a dataset and computing the value t of the test statistic. The p-value, on the other hand, represents the strength of the evidence the observed value t bears against H_0. But it does not specify all values of T that lead to rejection of H_0 at a given level α.

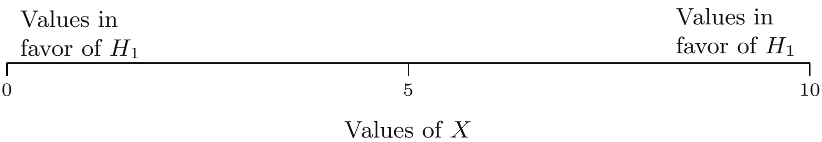
QUICK EXERCISE 26.2 In our freeway example, we have already computed the relevant tail probability to decide whether a person with recorded average speed $t = 124$ gets fined if we set the significance level at 0.05. Suppose the significance level is set at $\alpha = 0.01$ (we allow 1% of the drivers to get fined unjustly). Determine whether a person with recorded average speed $t = 124$ gets fined ($H_0 : \mu = 120$ is rejected). Furthermore, determine the critical region in this case.

Sometimes the critical region K can be constructed such that $\mathrm{P}(T \in K \mid H_0)$ is exactly equal to α, as in the freeway example. However, when the distribution of T is discrete, this is not always possible. This is illustrated by the next example.

After the introduction of the Euro, Polish mathematicians claimed that the Belgian 1 Euro coin is not a fair coin (see, for instance, the *New Scientist*, January 4, 2002). Suppose we put a 1 Euro coin to the test. We will throw it ten times and record X, the number of heads. Then X has a $Bin(10, p)$ distribution, where p denotes the probability of heads. We like to find out whether p differs from $1/2$. Therefore we test

$$H_0 : p = \frac{1}{2} \text{ (the coin is fair)} \quad \text{against} \quad H_1 : p \neq \frac{1}{2} \text{ (the coin is not fair)}.$$

We use X as the test statistic. When we set the significance level α at 0.05, for what values of X will we reject H_0 and conclude that the coin is not fair? Let us first find out what values of X are in favor of H_1. If $H_0 : p = 1/2$ is true, then $\mathrm{E}[X] = 10 \cdot \frac{1}{2} = 5$, so that values of X close to 5 are in favor H_0. Values close to 10 suggest that $p > 1/2$ and values close to 0 suggest that $p < 1/2$. Hence, both values close to 0 and values close to 10 are in favor of $H_1 : p \neq 1/2$.



This means that we will reject H_0 in favor of H_1 whenever $X \leq c_l$ or $X \geq c_u$. Therefore, the critical region is the set

$$K = \{0, 1, \ldots, c_l\} \cup \{c_u, \ldots, 9, 10\}.$$

The boundary values c_l and c_u are called *left* and *right* critical values. They must be chosen such that the critical region K is as large as possible and still satisfies

$$\mathrm{P}(X \in K \mid H_0) = \mathrm{P}\big(X \leq c_l \mid p = \tfrac{1}{2}\big) + \mathrm{P}\big(X \geq c_u \mid p = \tfrac{1}{2}\big) \leq 0.05.$$

Here $\mathrm{P}\big(X \geq c_u \mid p = \frac{1}{2}\big)$ denotes the probability $\mathrm{P}(X \geq c_u)$ computed with X having a $Bin(10, \frac{1}{2})$ distribution. Since we have no preference for rejecting H_0 for values close to 0 or close to 10, we divide 0.05 over the two sides, and we choose c_l as large as possible and c_u as small as possible such that

$$\mathrm{P}\big(X \leq c_l \mid p = \tfrac{1}{2}\big) \leq 0.025 \quad \text{and} \quad \mathrm{P}\big(X \geq c_u \mid p = \tfrac{1}{2}\big) \leq 0.025.$$

Table 26.1. Left tail probabilities of the $Bin(10, \frac{1}{2})$ distribution.

k	$P(X \leq k)$	k	$P(X \leq k)$
0	0.00098	6	0.82813
1	0.01074	7	0.94531
2	0.05469	8	0.98926
3	0.17188	9	0.99902
4	0.37696	10	1.00000
5	0.62305		

The left tail probabilities of the $Bin(10, \frac{1}{2})$ distribution are listed in Table 26.1. We immediately see that $c_l = 1$ is the largest value such that $P(X \leq c_l \mid p = 1/2) \leq 0.025$. Similarly, $c_u = 9$ is the smallest value such that $P(X \geq c_u \mid p = 1/2) \leq 0.025$. Indeed, when X has a $Bin(10, \frac{1}{2})$ distribution,

$$P(X \geq 9) = 1 - P(X \leq 8) = 1 - 0.98926 = 0.01074,$$
$$P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.94531 = 0.05469.$$

Hence, if we test $H_0 : p = 1/2$ against $H_1 : p \neq 1/2$ at level $\alpha = 0.05$, the critical region is the set $K = \{0, 1, 9, 10\}$. The corresponding type I error is

$$P(X \in K) = P(X \leq 1) + P(X \geq 9) = 0.01074 + 0.01074 = 0.02148,$$

which is smaller than the significance level. You may perform ten throws with your favorite coin and see whether the number of heads falls in the critical region.

Quick exercise 26.3 Recall the tank example where we tested $H_0 : N = 350$ against $H_1 : N < 350$ by means of the test statistic $T = \max X_i$. Suppose that we perform the test at level 0.05. Deduce the critical region K corresponding to level 0.05 from the left tail probabilities given here:

k	195	194	193	192	191
$P(T \leq k \mid H_0)$	0.0525	0.0511	0.0498	0.0485	0.0472

Is $P(T \in K \mid H_0) = 0.05$?

One- and two-tailed p-values

In the Euro coin example, we deviate from $H_0 : p = 1/2$ in *two* directions: values of X both far to the right and far to the left of 5 are evidence against H_0. Suppose that in ten throws with the 1 Euro coin we recorded x heads. What would the p-value be corresponding to x? The problem is that the direction in which values of X are *at least as extreme as* the observed value x depends on whether x lies to the right or to the left of 5.

At this point there are two natural solutions. One may report the appropriate left or right tail probability, which corresponds to the direction in which x deviates from H_0. For instance, if x lies to the right of 5, we compute $\mathrm{P}(X \geq x \mid H_0)$. This is called a *one-tailed p-value*. The disadvantage of one-tailed p-values is that they are somewhat misleading about how strong the evidence of the observed value x bears against H_0. In view of the relation between rejection on the basis of critical values or on the basis of a p-value, the one-tailed p-value should be compared to $\alpha/2$. On the other hand, since people are inclined to compare p-values with the significance level α itself, one could also double the one-tailed p-value and compare this with α. This double-tail probability is called a *two-tailed p-value*. It doesn't make much of a difference, as long as one *also* reports whether the reported p-value is one-tailed or two-tailed.

Let us illustrate things by means of the findings by the Polish mathematicians. They performed 250 throws with a Belgian 1 Euro coin and recorded heads 140 times (see also Exercise 24.2). The question is whether this provides strong enough evidence against $H_0 : p = 1/2$. The observed value 140 is to the right of 125, the value we would expect if H_0 is true. Hence the one-tailed p-value is $\mathrm{P}(X \geq 140)$, where now X has a $Bin(250, \frac{1}{2})$ distribution. By means of the normal approximation (see page 201), we find

$$\begin{aligned}\mathrm{P}(X \geq 140) &= \mathrm{P}\left(\frac{X-125}{\sqrt{\frac{1}{4}}\sqrt{250}} \geq \frac{140-125}{\sqrt{\frac{1}{4}}\sqrt{250}}\right)\\ &\approx \mathrm{P}(Z \geq 1.90) = 1-\Phi(1.90) = 0.0287.\end{aligned}$$

Therefore the two-tailed p-value is approximately 0.0574, which does not provide very strong evidence against H_0. In fact, the *exact* two-tailed p-value, computed by means of statistical software, is 0.066, which is even larger.

QUICK EXERCISE 26.4 In a Dutch newspaper (*De Telegraaf*, January 3, 2002) it was reported that the Polish mathematicians recorded heads 150 times. What are the one- and two-tailed probabilities is this case? Do they now have a case?

26.3 Type II error

As we have just seen, by setting a significance level α, we are able to control the probability of committing a type I error; it will at most be α. For instance, let us return to the freeway example and suppose that we adopt the decision rule to fine the driver for speeding if her average observed speed is at least 121.9, i.e.,

$$\text{reject } H_0 : \mu = 120 \text{ in favor of } H_1 : \mu > 120 \text{ whenever } T = \bar{X}_3 \geq 121.9.$$

From Section 26.1 we know that with this decision rule, the probability of a type I error is 0.05. What is the probability of committing a type II error? This corresponds to the percentage of drivers whose true speed *is* above 120 but who do not get fined because their recorded average speed is below 121.9.

For instance, suppose that a car passes at true speed $\mu = 125$. A type II error occurs when $T < 121.9$, and since $T = \bar{X}_3$ has an $N(125, 4/3)$ distribution, the probability that this happens is

$$\begin{aligned} \mathrm{P}(T < 121.9 \mid \mu = 125) &= \mathrm{P}\left(\frac{T - 125}{2/\sqrt{3}} < \frac{121.9 - 125}{2/\sqrt{3}}\right) \\ &= \Phi(-2.68) = 0.0036. \end{aligned}$$

This looks promising, but now consider a vehicle passing at true speed $\mu = 123$. The probability of committing a type II error in this case is

$$\begin{aligned} \mathrm{P}(T < 121.9 \mid \mu = 123) &= \mathrm{P}\left(\frac{T - 123}{2/\sqrt{3}} < \frac{121.9 - 123}{2/\sqrt{3}}\right) \\ &= \Phi(-0.95) = 0.1711. \end{aligned}$$

Hence 17.11% of all drivers that pass at speed $\mu = 123$ will not get fined. In Figure 26.3 the last situation is illustrated. The curve on the left represents the probability density of the $N(120, 4/3)$ distribution, which is the distribution of T under the null hypothesis. The shaded area on the right of 121.9 represents the probability of committing a type I error

$$\mathrm{P}(T \geq 121.9 \mid \mu = 120) = 0.05.$$

The curve on the right is the probability density of the $N(123, 4/3)$ distribution, which is the distribution of T under the alternative $\mu = 123$. The shaded area on the left of 121.9 represents the probability of a type II error

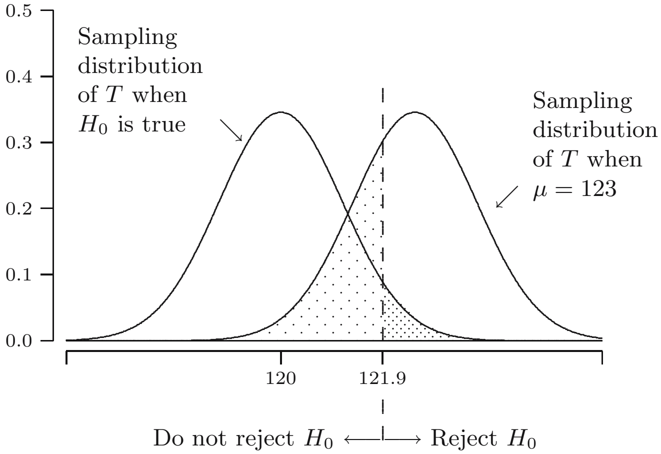


Fig. 26.3. Type I and type II errors in the freeway example.

$$\mathrm{P}(T < 121.9 \mid \mu = 123) = 0.1711.$$

Shifting μ further to the right will result in a smaller probability of a type II error. However, shifting μ toward the value 120 leads to a larger probability of a type II error. In fact it can be arbitrarily close to 0.95.

The previous example illustrates that the probability of committing a type II error depends on the actual value of μ in the alternative hypothesis $H_1 : \mu > 120$. The closer μ is to 120, the higher the probability of a type II error will be. In contrast with the probability of a type I error, which is always at most α, the probability of a type II error may be arbitrarily close to $1 - \alpha$. This is illustrated in the next quick exercise.

Quick exercise 26.5 What is the probability of a type II error in the freeway example if $\mu = 120.1$?

26.4 Relation with confidence intervals

When testing $H_0 : \mu = 120$ against $H_1 : \mu > 120$ at level 0.05 in the freeway example, the critical value was obtained by the formula

$$c_{0.05} = 120 + 1.645 \cdot \frac{2}{\sqrt{3}}.$$

On the other hand, using that $\bar{X}_3$ has an $N(\mu, 4/3)$ distribution, a 95% lower confidence bound for μ in this case can be derived from

$$l_n = \bar{x}_3 - 1.645 \cdot \frac{2}{\sqrt{3}}.$$

Although, at first sight, testing hypotheses and constructing confidence intervals seem to be two separate statistical procedures, they are in fact intimately related. In the freeway example, observe that for a given dataset x_1, x_2, x_3,

$$\begin{aligned}
&\text{we reject } H_0 : \mu = 120 \text{ in favor of } H_1 : \mu > 120 \text{ at level } 0.05\\
\Leftrightarrow\quad & \bar{x}_3 \geq 120 + 1.645 \cdot \frac{2}{\sqrt{3}}\\
\Leftrightarrow\quad & \bar{x}_3 - 1.645 \cdot \frac{2}{\sqrt{3}} \geq 120\\
\Leftrightarrow\quad & 120 \text{ is not in the 95\% one-sided confidence interval for } \mu.
\end{aligned}$$

This is not a coincidence. In general, the following applies. Suppose that for some parameter θ we test $H_0 : \theta = \theta_0$. Then

we reject $H_0 : \theta = \theta_0$ in favor of $H_1 : \theta > \theta_0$ at level α

if and only if

θ_0 is not in the $100(1 - \alpha)\%$ *one-sided* confidence interval for θ.

The same relation holds for testing against $H_1 : \theta < \theta_0$, and a similar relation holds between testing against $H_1 : \theta \neq \theta_0$ and two-sided confidence intervals:

we reject $H_0 : \theta = \theta_0$ in favor of $H_1 : \theta_0 \neq \theta_0$ at level α
if and only if
θ_0 is not in the $100(1-\alpha)\%$ *two-sided* confidence region for θ.

In fact, one could use these facts to define the $100(1-\alpha)\%$ *confidence region* for a parameter θ as the set of values θ_0 for which the null hypothesis $H_0 : \theta = \theta_0$ is *not rejected* at level α.

It should be emphasized that these relations *only* hold if the random variable that is used to construct the confidence interval relates appropriately to the test statistic. For instance, the preceding relations do not hold if on the one hand, we construct a confidence interval for the parameter μ of an $N(\mu, \sigma^2)$ distribution by means of the studentized mean $(\bar{X}_n - \mu)/(S_n/\sqrt{n})$, and on the other hand, use the sample median Med_n to test a null hypothesis for μ.

26.5 Solutions to the quick exercises

26.1 In the first situation, we reject at significance level $\alpha = 0.05$, which means that the probability of committing a type I error is at most 0.05. This does not necessarily mean that this probability will also be less than or equal to 0.01. Therefore with this information we cannot know whether we also reject at level $\alpha = 0.01$. In the reversed situation, if we reject at level $\alpha = 0.01$, then the probability of committing a type I error is at most 0.01, and is therefore also smaller than 0.05. This means that we also reject at level $\alpha = 0.05$.

26.2 To decide whether we should reject $H_0 : \mu = 120$ at level 0.01, we could compute $\mathrm{P}(T \geq 124 \mid H_0)$ and compare this with 0.01. We have already seen that $\mathrm{P}(T \geq 124 \mid H_0) = 0.0003$. This is (much) smaller than the significance level $\alpha = 0.01$, so we should reject.

The critical region is $K = [c, \infty)$, where we must solve c from

$$\mathrm{P}\left(Z \geq \frac{c - 120}{2/\sqrt{3}}\right) = 0.01.$$

Since $z_{0.01} = 2.326$, this means that $c = 120 + 2.326 \cdot (2/\sqrt{3}) = 122.7$.

26.3 The critical region is of the form $K = \{5, 6, \ldots, c\}$, where the critical value c is the largest value, for which $\mathrm{P}(T \leq c \mid H_0)$ is still less than or equal to 0.05. From the table we immediately see that $c = 193$ and that $\mathrm{P}(T \in K \mid H_0) = \mathrm{P}(T \leq 193 \mid H_0) = 0.0498$, which is not equal to 0.05.

26.4 By means of the normal approximation, for the one-tailed p-value we find

$$\begin{aligned} \mathrm{P}(X \geq 150) &= \mathrm{P}\left(\frac{X - 125}{\sqrt{\frac{1}{4}}\sqrt{250}} \geq \frac{150 - 125}{\sqrt{\frac{1}{4}}\sqrt{250}} \right) \\ &= \mathrm{P}(Z_n \geq 3.16) \approx 1 - \Phi(3.16) = 0.0008. \end{aligned}$$

The two-tailed p-value is 0.0016. This is a lot smaller than the two-tailed p-value 0.0574, corresponding to 140 heads. It seems that with 150 heads the mathematicians would have a case; the Belgian Euro coin would then appear not to be fair.

26.5 The probability of a type II error is

$$\begin{aligned} \mathrm{P}(T < 121.9 \mid \mu = 120.1) &= \mathrm{P}\left(\frac{T - 120.1}{2/\sqrt{3}} < \frac{121.9 - 120.1}{2/\sqrt{3}} \right) \\ &= \Phi(1.56) = 0.9406. \end{aligned}$$

26.6 Exercises

26.1 Polygraphs that are used in criminal investigations are supposed to indicate whether a person is lying or telling the truth. However the procedure is not infallible, as is illustrated by the following example. An experienced polygraph examiner was asked to make an overall judgment for each of a total 280 records, of which 140 were from guilty suspects and 140 from innocent suspects. The results are listed in Table 26.2. We view each judgment as a problem of hypothesis testing, with the null hypothesis corresponding to "suspect is innocent" and the alternative hypothesis to "suspect is guilty." Estimate the probabilities of a type I error and a type II error that apply to this polygraph method on the basis of Table 26.2.

26.2 Consider the testing problem in Exercise 25.11. Compute the probability of committing a type II error if the true value of μ is 1.

26.3 ⊡ One generates a number x from a uniform distribution on the interval $[0, \theta]$. One decides to test $H_0 : \theta = 2$ against $H_1 : \theta \neq 2$ by rejecting H_0 if $x \leq 0.1$ or $x \geq 1.9$.

- **a.** Compute the probability of committing a type I error.
- **b.** Compute the probability of committing a type II error if the true value of θ is 2.5.

26.4 To investigate the hypothesis that a horse's chances of winning an eight-horse race on a circular track are affected by its position in the starting lineup,

Table 26.2. Examiners and suspects.

		Suspect's true status	
		Innocent	Guilty
Examiner's assesment	Acquitted	131	15
	Convicted	9	125

Source: F.S. Horvath and J.E. Reid. The reliability of polygraph examiner diagnosis of truth and deception. *Journal of Criminal Law, Criminology, and Police Science*, 62(2):276–281, 1971.

the starting position of each of 144 winners was recorded ([30]). It turned out that 29 of these winners had starting position one (closest to the rail on the inside track). We model the number of winners with starting position one by a random variable T with a $Bin(144, p)$ distribution. We test the hypothesis $H_0 : p = 1/8$ against $H_1 : p > 1/8$ at level $\alpha = 0.01$ with T as test statistic.

a. Argue whether the test procedure involves a right critical value, a left critical value, or both.

b. Use the normal approximation to compute the critical value(s) corresponding to $\alpha = 0.01$, determine the critical region, and report your conclusion about the null hypothesis.

26.5 ⊞ Recall Exercises 23.5 and 24.8 about the 1500 m speed-skating results in the 2002 Winter Olympic Games. The number of races won by skaters starting in the outer lane is modeled by a random variable X with a $Bin(23, p)$ distribution. The question of whether there is an outer lane advantage was investigated in Exercise 24.8 by means of constructing confidence intervals using the normal approximation. In this exercise we examine this question by testing the null hypothesis $H_0 : p = 1/2$ against $H_1 : p > 1/2$ using X as the test statistic. The distribution of X under H_0 is given in Table 26.3. Out of 23 completed races, 15 were won by skaters starting in the outer lane.

a. Compute the p-value corresponding to $x = 15$ and report your conclusion if we perform the test at level 0.05. Does your conclusion agree with the confidence interval you found for p in Exercise 24.8 **b**?

b. Determine the critical region corresponding to significance level $\alpha = 0.05$.

c. Compute the probability of committing a type I error if we base our decision rule on the critical region determined in **b**.

Table 26.3. Left tail probabilities for the $Bin(23, \frac{1}{2})$ distribution.

k	$P(X \leq k)$	k	$P(X \leq k)$	k	$P(X \leq k)$
0	0.0000	8	0.1050	16	0.9827
1	0.0000	9	0.2024	17	0.9947
2	0.0000	10	0.3388	18	0.9987
3	0.0002	11	0.5000	19	0.9998
4	0.0013	12	0.6612	20	1.0000
5	0.0053	13	0.7976	21	1.0000
6	0.0173	14	0.8950	22	1.0000
7	0.0466	15	0.9534	23	1.0000

d. Use the normal approximation to determine the probability of committing a type II error for the case $p = 0.6$, if we base our decision rule on the critical region determined in **b**.

26.6 ⊡ Consider Exercises 25.2 and 25.7. One decides to test $H_0 : \mu = 1472$ against $H_1 : \mu > 1472$ at level $\alpha = 0.05$ on the basis of the recorded value 1718 of the test statistic T.

a. Argue whether the test procedure involves a right critical value, a left critical value, or both.

b. Use the fact that the distribution of T can be approximated by an $N(\mu, \mu)$ distribution to determine the critical value(s) and the critical region, and report your conclusion about the null hypothesis.

26.7 A random sample X_1, X_2 is drawn from a uniform distribution on the interval $[0, \theta]$. We wish to test $H_0 : \theta = 1$ against $H_1 : \theta < 1$ by rejecting if $X_1 + X_2 \leq c$. Find the value of c and the critical region that correspond to a level of significance 0.05.
Hint: use Exercise 11.5.

26.8 ⊞ This exercise is meant to illustrate that the shape of the critical region is not necessarily similar to the type of alternative hypothesis. The type of alternative hypothesis *and* the test statistic used determine the shape of the critical region.

Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from an $Exp(\lambda)$ distribution, and we test $H_0 : \lambda = 1$ with test statistics $T = \bar{X}_n$ and $T' = e^{-\bar{X}_n}$.

a. Suppose we test the null hypothesis against $H_1 : \lambda > 1$. Determine for both test procedures whether they involve a right critical value, a left critical value, or both.

b. Same question as in part **a**, but now test against $H_1 : \lambda \neq 1$.

26.9 ⊞ Similar to Exercise 26.8, but with a random sample $X_1, X_2, \ldots, X_n$ from an $N(\mu, 1)$ distribution. We test $H_0 : \mu = 0$ with test statistics $T = (\bar{X}_n)^2$ and $T' = 1/\bar{X}_n$.

a. Suppose that we test the null hypothesis against $H_1 : \mu \neq 0$. Determine the shape of the critical region for both test procedures.

b. Same question as in part **a**, but now test against $H_1 : \mu > 0$.