In this chapter, we are interested in the question of power and sample size for comparing two samples. The samples may come from populations with normal, binomial or Poisson densities and our estimates of power and sample size refer to differences between means, proportions and rates.

Let us summarize the issues involved with power and sample size. In planning a two-sample study, we must guard against two types of errors. The first is Type I error. It refers to declaring the difference in, for example, proportions significant when in fact it is not. To guard against this error, we set \( \alpha \) to be as small as we can tolerate—usually 0.1, 0.05 or 0.01. By increasing sample size, we can also achieve the desired significance, no matter how small the difference is between two proportions. So we need to specify a difference as large as we deem detectable. The second is Type II error. Here we declare the difference between two population parameters (means, proportions or intensities) as significant while in fact it is not. So after we specify the minimum difference that is important to be detected, we need to specify the probability of detecting this difference. This probability, denoted by \( 1 - \beta \), determines the power of the test. Recall that \( \beta \) is the probability of Type II error. To compute a necessary sample size, we specify the minimum detectable difference between the parameters of interest, the desired significance and the desired power.

### 13.1 Two means from normal populations

Here we discuss how to obtain the power to distinguish the difference between the means of two populations based on two samples. We shall also see how to obtain sample sizes necessary to distinguish between the means with a given difference, significance and power.

#### 13.1.1 Power

The hypotheses to be tested are \( H_0 : \mu = 0 \) vs. one of the usual three alternatives for a specified \( \alpha \). Here \( \mu := \mu_2 - \mu_1 \). To obtain the power to distinguish between two
means from normal populations based on two samples from these populations, we must specify a value for the alternative difference between the means, denoted by $\mu_A$. Let

$$SE := \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

where $n_i$ and $S_i^2$ ($i = 1, 2$) are the respective means and variances of the two samples. Denote by $P(Z < z)$ the probability that the rv $Z$ takes on values less than $z$ where $Z$ is from a standard normal distribution. For the hypotheses $H_0 : \mu = 0$ vs. $H_A : \mu \neq 0$ and for a given alternative $\mu_A$, the two-sided power is given by

$$1 - \beta = P\left( Z < \frac{\mu_A}{SE} - z_{1-\alpha/2} \right) + P\left( Z < -\frac{\mu_A}{SE} - z_{\alpha/2} \right)$$

$$= \text{pnorm}(\mu.A/SE - \text{qnorm}(1 - \alpha/2)) + \text{pnorm}(-\mu.A/SE - \text{qnorm}(\alpha/2)). \tag{13.1}$$

For $H_A : \mu > \mu_2 - \mu_1$, the power is given by

$$1 - \beta = P\left( Z < \frac{\mu_A}{SE} - z_{1-\alpha} \right)$$

$$= \text{pnorm}(\mu.A/SE - \text{qnorm}(1 - \alpha)) \tag{13.2}$$

and for $H_A : \mu < \mu_2 - \mu_1$, the power is given by

$$1 - \beta = P\left( Z < \frac{\mu_A}{SE} - z_{\alpha} \right)$$

$$= \text{pnorm}(\mu.A/SE - \text{qnorm}(\alpha)) \tag{13.3}.$$ 

**Example 13.1.** Consider the capital punishment data first introduced in Example 2.12. To examine differences of age at sentencing between blacks and whites, we sample the data with $n_1 = 35$, $n_2 = 40$ and find that $X_1 = 29.0$, $X_2 = 27.6$, $\sigma_1^2 = 64.8$ and $\sigma_2^2 = 61.1$ respectively. The $p$-value for the difference in the means is $0.112 > 0.025$ and we do not reject the hypothesis that the mean age at sentencing is equal for whites and blacks. How powerful is our ability to distinguish between these two means if in fact the difference between the true (population) means is $|\mu| = |27.6 - 29| = 1.4$? Before answering this question, let us first examine the power profiles according to Equations (13.1), (13.2) and (13.3):

```r
source('power-normal.R')
alpha <- 0.05 ; mu.0 <- 0 ; mu.1 <- 29 ; mu.2 <- 27.6
mu.A <- seq(-10, 10, length = 201) ; V.1 <- 64.8
V.2 <- 61.1 ; n.1 <- 35 ; n.2 <- 40 ; k <- n.2 / n.1
par(mfrow = c(1, 3))
alt <- c('two.sided', 'greater', 'less')
for (i in 1 : 3){
  if(i == 1) ylab = 'power' else ylab = ''
  p <- power.normal(mu.A = mu.A, mu.0 = mu.0, n.1 = n.1,
```
Let us explain the code. In line 1, we execute a script in which the function `power.normal()` resides. The code for this function resides in `power-normal.R` at the book website. In lines 3–5 we specify the data. We set \( \mu_0 := \mu_2 - \mu_1 = 0 \). In other words, we are looking for the power to distinguish between two populations’ means under the null hypothesis that the means are not different at \( \alpha = 0.05 \). In line 4, we set the alternative difference between the population means as a vector (we wish to examine the power profile). In lines 4 and 5 we specify the variances, sample sizes and the ratio of the sample sizes. Because we wish to plot the profiles for the three \( H_A \), we open a window ready to accept a matrix of plots with one row and three columns (line 7). We store in `alt` the three power types we wish to examine and plot.

In lines 9–16 we call `power.normal()` and plot the results. The function can calculate power for a single or two samples from a normal density. For two samples, the function requires the arguments as shown. Here \( \text{S.1} \) and \( \text{S.2} \) denote the standard deviations of the samples and `alt` denotes the alternative for which we wish to determine the power. The resulting power profiles are shown in Figure 13.1. To obtain the power for \( |\mu_A| = 1.4 \) under \( H_0: \mu = 0 \), we simply call

```r
> mu.A <- 1.4
> p <- power.normal(mu.A, mu.0, n.1, n.2, S.1 = sqrt(V.1),
+   S.2 = sqrt(V.2), alt = 'two.sided')
> p$pwr
                 mu.A    power
1 1.4 0.1186401
```

Figure 13.1  Power profiles for distinguishing between the age of sentencing to death of blacks and white inmates in the U.S.
Thus, our ability to distinguish a difference of 1.4 years between the mean ages at sentencing to death of whites and blacks is negligible; i.e. $1 - \beta \approx 0.189$.

13.1.2 Sample size

To compare two means of two samples from normal populations with

$$H_0: \mu = \mu_0 \quad \text{vs.} \quad H_A: \mu \neq \mu_0$$

with significance level $\alpha$ and power $1 - \beta$ we need to specify the smallest detectable difference. Recall that $\mu := \mu_2 - \mu_1$. Also recall that because we are dealing with large samples, we use $\sigma_1 \approx S_1$ and $\sigma_2 \approx S_2$ where $S_1$ and $S_2$ are the sample-based standard deviations of $X_1$ and $X_2$. If we have no idea about the population standard deviations, we may use the range of the data divided by 4 to estimate the variance and then the standard deviations. Let $n$ be the sample size of each of the two samples. Then, for the two-tailed estimate, the smallest sample size we need is

$$n = \frac{(\sigma_1^2 + \sigma_2^2) (z_{1-\alpha/2} + z_{1-\beta})^2}{\mu^2}.$$

Often, because of cost or other concerns, we anticipate that $n_2$ will be larger than $n_1$ by a factor $k$; i.e. $n_2 = k \times n_1$. In such cases, we estimate the needed sample size with

$$n_1 = \frac{(\sigma_1^2 + \sigma_2^2/k) (z_{1-\alpha/2} + z_{1-\beta})^2}{\mu^2},$$

$$n_2 = \frac{(k\sigma_1^2 + \sigma_2^2) (z_{1-\alpha/2} + z_{1-\beta})^2}{\mu^2}.$$

For one-tailed estimates, replace $z_{1-\alpha/2}$ above with $z_{1-\alpha}$.

Example 13.2. We continue with the capital punishment data (Example 12.1). From the samples we had, we specify $\sigma_1^2 \approx 64.8$ for whites’ age at sentencing to death and $\sigma_2^2 \approx 61.1$ for blacks. We wish to calculate the sample sizes that we need to obtain a significant difference at $\alpha = 0.05$ with $1 - \beta$ between 0.6 and 0.9 for detectable differences between $-5$ and 5 years of age. The following script accomplishes the task.

```r
alpha <- 0.05 ; mu.0 <- 0 ; V.1 <- 64.8 ; V.2 <- 61.1
mu <- c(seq(-8, 8, length = 161))
pwr <- seq(0.6, 0.9, length = 30)
s <- sample.size.normal(mu, S.1 = sqrt(V.1),
                        S.2 = sqrt(V.2), power = pwr, alt = alt[i])
sm <- matrix(s$sSize$n.1, ncol = length(mu),
            nrow = length(pwr), byrow = TRUE)
```
The function that computes sample size for one or two samples is called `sample.size.normal()`. It is available from the book website in the link for the file `sample-size-normal.R`. In lines 1 to 3 we prepare the data. In lines 5 and 6 we call the function with both power and $\mu$ vectors (their values are set in lines 2 and 3). `sample.size.normal()` returns a list with the data and the output. The latter is stored in the list as a data frame. Here are some of its lines:

```r
> head(s$size)
  mu  power n.1 n.2
1 -8.0 0.6  8  8
2 -7.9 0.6  8  8
3 -7.8 0.6  8  8
4 -7.7 0.6  8  8
5 -7.6 0.6  9  9
6 -7.5 0.6  9  9
```

In line 8 we remove values of $n_1$ and $n_2$ from the results for those absolute values of $\mu$ that are too close to zero because the sample sizes for these values are either too large, or because a detectable difference between $\pm 2$ years is not important.

To prepare the function output for a 3D plot, we create a matrix from the values of $n_1$. This matrix has as many columns as the length of the vector $\mu$ and as many rows as the length of $\text{pwr}$ (the former contains the values of the differences between $\mu_2$ and $\mu_1$ and the latter the values of the power).

Finally, in lines 12 to 14, we call the R function `persp()` (see Figure 13.2). To obtain the sample size for $\alpha = 0.05$, for a detectable age difference of 2.5 years between the mean ages of blacks and whites at the time of sentencing and for $1 - \beta = 0.8$, we first set the condition for extraction of the results from the `s$size` data frame:

```r
> condition <- round(s$size$mu,2) == 2.50 & +   round(s$size$power,2) == 0.80
```

and then

```r
> s$size[condition, ]
   mu  power n.1 n.2
3165 2.5 0.7965517 124 124
```

In other words, to detect the desired difference in age with the desired power, we need a sample of 124 whites and blacks. Let us see if this indeed is the case. In sampling the data from the population, we `set.seed()` to 10 and $n_1 = n_2 = 124$. This gives

```
whites    blacks
mean 32.7   27.9
variance 98.1  51.5
sample size 124.0 124.0
```
and a \( p \)-value of

\[
> 1 - \text{pnorm(X.bar.1 - X.bar.2, 0, SE)}
\]

[1] 7.488974e-06

We thus conclude that if 2.5 years of age-difference (of blacks and whites at the time of sentencing to death) indicates, for example, prejudice against blacks (in sentencing young to death), then a sample of 124 will suffice to detect this difference. \( \square \)

Example 13.2 illustrates an extremely important point. One of the most frequent criticisms of the abuse of statistics is this: You can always establish a significant difference if you use a large enough sample. We know by now that this criticism is valid because the standard deviation of the sampling density (the standard error) decreases as the sample size increases. So if you have a large enough population, you can always establish a significant difference by increasing your sample size (recall our bigot in Example 10.14). \textit{However}, if common sense dictates that the smallest detectable difference \( \mu := \mu_2 - \mu_1 \) makes sense, then you can calculate the sample size needed to detect this difference and thus avoid abusing statistics. In the case of our capital punishment example, we decide (for whatever reason) that a detectable difference in mean age of sentencing of at least 2.5 years between blacks and white may be practically important. Thus, any sample larger than 124 will amount to “forcing the issue.”

### 13.2 Two proportions

Here, we follow the same sequence as we did in Section 13.1. Unlike the power obtained from comparing two means, we usually do not have repeated experiments. That is, we must distinguish between \( n_i \) representing repetitions (as in Section 13.1) and between
two experiments, one with $n_1$ trials and $n_1S$ successes and the other with $n_2$ trials and $n_2S$ success.

13.2.1 Power

We are interested in the power to distinguish between proportions from two populations. To obtain it, we must specify the level of significance, the difference between $\pi_1$ and $\pi_2$ that is important to detect and whether we are testing for one- as opposed to two-tailed hypotheses.

The hypotheses to be tested are $H_0: \pi = 0$ vs. one of the usual three alternatives for a specified $\alpha$. Here $\pi := \pi_2 - \pi_1$. As usual, $\pi_1$ and $\pi_2$ are the probabilities of success in the respective populations. These are estimated with $\pi_i \approx p_i = n_iS / n_i$, $i = 1, 2$.

Under the assumption that $\pi_1 = \pi_2$, we have that the mean of $\pi$, denoted by $\bar{\pi}$ and its standard error, $SE$, are

$$\bar{\pi} = \frac{n_1\pi_1 + n_2\pi_2}{n_1 + n_2}, \quad SE = \sqrt{\bar{\pi}(1 - \bar{\pi}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.$$  

We use the samples’ proportions of success, $p_1$ and $p_2$, to estimate $\pi_1$ and $\pi_2$. Consequently, the standard error of the sampling distribution of $\pi_2 - \pi_1$ is given by

$$SE \approx \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}.$$  

For the two-sided power (i.e. $H_A: \pi \neq \pi_2 - \pi_1$), the power is given by

$$1 - \beta = 1 - P \left( Z < \frac{z_{1-\alpha/2} SE - |\pi|}{SE} \right) + P \left( Z < \frac{-z_{1-\alpha/2} SE - |\pi|}{SE} \right)$$

where $P(Z < z)$ is the probability (area) under the standard normal density that $Z < z$. For $H_A: \pi_2 > \pi_1$, the one-sided “greater than” power is given by

$$1 - P \left( Z < \frac{z_{1-\alpha} SE - |\pi|}{SE} \right)$$

and for the “less than” $H_A: \pi_2 < \pi_1$, the power is given by

$$P \left( Z < \frac{-z_{1-\alpha} SE - |\pi|}{SE} \right).$$

Example 13.3. Two groups of 40 patients each were selected for a study of the effectiveness of flu shots. Members of the treatment group received a flu shot. Members of the control group received a saline shot. The medical history of both groups was followed for the duration of the flu season. Of the control group, 15 suffered from flu symptoms at least once. Of the treatment group, 10 did. We wish to answer the following:

1. Was the treatment effective?
2. If not, what is the probability that we accept the hypothesis that the treatment was not effective in preventing flu while in fact it was (i.e. type II error, $\beta$)?
3. What should have been the number of people in the treatment group that did not suffer from flu symptoms for a power of 0.8; i.e. for a power that will guarantee a small (0.2) type II error?

To answer these questions, we first set the notation:

Treatment: \( n_1 = 40 \), \( n_{1S} = 10 \), \( p_1 = \frac{n_{1S}}{n_1} = 0.25 \).

Control: \( n_2 = 40 \), \( n_{2S} = 15 \), \( p_2 = \frac{n_{2S}}{n_2} = 0.375 \).

Hypotheses: \( H_0: \pi_1 = \pi_2 \), \( H_A: \pi_1 < \pi_2 \), \( k = \frac{n_2}{n_1} = 1 \),
\[ \pi_1 \approx p_1 \, , \, \pi_2 \approx p_2 \, , \, \bar{p} = \frac{n_{1S} + n_{2S}}{n_1 + n_2} = 0.3125 \, . \]

Regarding the first question, we have

```r
> prop.test(c(10, 15), c(40, 40), alternative = 'g')
2-sample test for equality of proportions with continuity correction
data: c(10, 15) out of c(40, 40)
X-squared = 0.9309, df = 1, p-value = 0.8327
alternative hypothesis: greater
95 percent confidence interval:
-0.3189231 1.0000000
sample estimates:
prop 1 prop 2
0.250 0.375
```

and we do not reject the null hypothesis. Therefore, we conclude that flu shots were not effective.

To answer the second question, we set the data and call `bp()` (for binomial power), available in `bp.R`, at the book’s site, with a one sided test:

```r
> source('bp.R')
> n <- c(40, 40) ; n.S <- c(10, 15) ; p <- n.S / n
> Power <- bp(p[1], p[2], n1 = n[1], n2 = n[2],
+ alt = 'greater')
> print(c(beta = 1 - as.vector(Power)))
beta
0.6710638
```

Therefore, the type II error is approximately 0.671. In other words, the probability that we accept the hypothesis that the treatment was not effective in preventing flu while in fact it was is 0.671—not a good state of affairs because we may deny effective treatment.

To answer the third question, we do:

```r
> pi.A <- seq(0, p[2], length = 201)
> Power <- bp(pi.A, p[2], n1 = n[1], n2 = n[2],
+ alt = 'greater')
> plot(pi.A, Power, xlab = expression(pi[A]), type = 'l')
```
Two proportions 409

Figure 13.3 One sided power profile for $\pi_2 = 0.375 > \pi_A$ between 0 and $\pi_2$.

(see Figure 13.3). Thus we find

```r
> c(pi.A = pi.A[72], bp(pi.A[72], p[2], n1 = n[1],
+   n2 = n[2], alt = 'greater'))
pi.A Power
0.1331250 0.8090285
> floor(pi.A[72] * n[1])
[1] 5
```

In other words, in the current experiment, we needed no more than five people from the experiment group contracting the flu to obtain a power of approximately 0.8. Such power presents a balance between the probability of denying effective treatment (0.2) and the probability of providing flu shots while they are not effective (0.05). Under such conditions, it might be reasonable to select $\alpha = 0.1$ for then we will decrease the probability of denying effective treatment.

13.2.2 Sample size

Here we are interested in determining the sample size needed to distinguish between two proportions with a particular power and level of significance.

Let $\rho := n_2 / n_1$. Under the null ($\pi_2 = \pi_1$) and alternative ($\pi_2 \neq \pi_1$) hypotheses, we first obtain the pooled proportion

$$\overline{\pi} := \frac{(\pi_1 + \rho \pi_2)}{1 + \rho}.$$

Next, the standard deviations under the null and under the alternative, where for the alternative we specify $\pi_A := |\pi_2 - \pi_1|$, are

$$\sigma_0 = \sqrt{\overline{\pi} (1 - \overline{\pi}) \left(1 + \frac{1}{\rho}\right)},$$

$$\sigma_A = \sqrt{\pi_1 (1 - \pi_1) + \frac{\pi_2 (1 - \pi_2)}{\rho}}.$$
Then, the two-sided sample size is obtained from

\[ n' = \left[ \frac{z_{1-\alpha/2} \times \sigma_0 + z_{1-\beta} \times \sigma_A}{\pi_A} \right]^2, \]

\[ n_2 = \text{largest integer closest to } \rho \times n', \]

\[ n_1 = \text{largest integer closest to } n'/\rho. \]

For a one-sided test, use \( z_{1-\alpha} \).

Often, cost and other considerations dictate that the sample sizes should be different. This can be achieved by using appropriate values of \( \rho \).

**Example 13.4.** Continuing with Example 13.3, we wish to determine the sample sizes that are necessary to establish a difference of \( 0.375 - 0.25 \) between the proportion that got sick in the control and treatment groups. We use the standard values of \( \alpha = 0.05 \) and \( 1 - \beta = 0.8 \) and the same fraction of the total sample allocated to both groups. Then

\[
> \text{library(Hmisc)}
> \text{ceiling(bsamsize(p.1, p.2))}
> n1 \ n2
435 435
\]

In other words, we need 870 people to achieve the desired significance. Suppose that it is twice as expensive to follow members of the treatment group compared to the control group e.g. following a member of the treatment group costs $100 and following a member of the control group costs $50. Then, our desired fraction of allocation to the treatment group is \( 1/3 \) and

\[
> \text{ceiling(bsamsize(p.1, p.2, fraction = 1/3))}
> n1 \ n2
312 624
\]

The cost for the treatment group is \( 312 \times 100 = 31200 \). The cost for the control group is \( 624 \times 50 = 31200 \) for a total cost of $62 400. Here we need more people (936) compared to equal sample sizes (870). We may wish to investigate the possibility of allocating the 936 people to both groups in a way that will maximize the power we can achieve. Then

\[
> \text{ba <- ballocation(p.1, p.2, 936)}
> \text{as.vector(c(936 * ba[4], ba[4]))}
[1] 442.2857658 0.4725275
\]

Thus, we conclude that instead of allocating 312, we may allocate 443 (of the 936) to the treatment. This will maximize the power we expect to achieve at a cost of $68 900 (compared to $62 450 when no power-maximizing is considered). \( \square \)

### 13.3 Two rates

Let \( t' \) denote the time from the occurrence of the last event. Denote by \( P(X < 1 | t') \) the probability that no event occurred by \( t' \). As \( t' \) increases, this probability decreases
because the more time passes since the time of last event, the smaller the probability
that the event does not occur. It can be shown that if \( X \) is Poisson with \( \lambda \), then
\[
P(X < 1|t') = e^{-\lambda t'}.
\]
Therefore,
\[
P(X \geq 1|t') = 1 - e^{-\lambda t'}.
\]
For a sample of size \( n \), the expected number of events is then
\[
m := nP(X \geq 1|t') = n \left( 1 - e^{-\lambda t'} \right).
\]
For two Poisson populations we have \( \lambda_1, \lambda_2, t_1', t_2', n_1, n_2, m_1 \) and \( m_2 \). Denote by \( \pi \) the probability of an event from \( n_1 \). Suppose we observe \( n_1 \) for \( t_1 \) time units and \( n_2 \) for \( t_2 \) time units. Then
\[
T_1 = n_1 t_1, \quad T_2 = n_2 t_2, \quad \pi = \frac{\lambda_1 T_1}{\lambda_1 T_1 + \lambda_2 T_2}.
\]  
(13.4)

Events are independent. Therefore, the number of events from \( n_1 \) is binomial with parameters \( \pi \) and \( m_1 + m_2 \). During the time we follow subjects (\( t_1' \) and \( t_2' \)), we expect that
\[
m = m_1 + m_2 = n_1 \left( 1 - e^{-\lambda_1 t_1'} \right) + n_2 \left( 1 - e^{-\lambda_2 t_2'} \right)
\]
events will occur.

To proceed, we define \( \rho := \lambda_1 / \lambda_2 \). Then dividing the numerator and the denominator of the expression for \( \pi \) in (13.4) by \( \lambda_2 \), we obtain
\[
\pi = \frac{\lambda_1 T_1}{\lambda_1 T_1 + \lambda_2 T_2} = \frac{\lambda_1 T_1 / \lambda_2}{T_1 \rho / T_2}.
\]

We wish to test the hypothesis that the rates \( \lambda_1 \) and \( \lambda_2 \) are equal. So we set
\[
H_0 : \rho = 1 \quad \text{vs.} \quad \rho > 1
\]
which is equivalent to
\[
H_0 : \pi = \frac{T_1}{T_1 + T_2} \quad \text{vs.} \quad \pi > \frac{T_1}{T_1 + T_2}.
\]
To simplify the notation, we let \( \pi_0 := T_1 / (T_1 + T_2) \) and \( \pi_A > \pi_0 \), where \( \pi_A \) is specified. So equivalent to \( H_0 : \rho = 1 \) we have \( H_0 : \pi = \pi_0 \), with the alternative specified. Thus, similar to the development in Section 11.1.2, for \( \pi_A > \pi_0 \), we obtain
\[
\text{power} = P \left( Z \leq \frac{(\pi_A - \pi_0) \sqrt{m} - z_{1-\alpha} \sqrt{V_0}}{\sqrt{V_A}} \right)
\]  
(13.5)

where
\[
V_0 := \pi_0 (1 - \pi_0) \quad \text{and} \quad V_A := \pi_A (1 - \pi_A).
\]
For \( \pi_A < \pi_0 \), we use
\[
\text{power} = P \left( Z \leq \frac{(\pi_0 - \pi_A) \sqrt{m} - z_{1-\alpha} \sqrt{V_0}}{\sqrt{V_A}} \right). \tag{13.6}
\]

To obtain two-sided power, replace \( z_{1-\alpha} \) by \( z_{1-\alpha/2} \) and sum the right hand sides of equations (13.5) and (13.6).

Example 13.5. The incidence rate of a genetic mutation in population 1 is 375 per 100,000 in one year. In population 2 it is 300 per 100,000 in one year. We take a sample of 5000 from each population. What is the power of distinguishing between \( \lambda_1 = 375 \times 10^{-5} \) and \( \lambda_2 = 300 \times 10^{-5} \) at \( \alpha = 0.05 \)?

The expected number of incidences in \( t' = 5 \) years are
\[
m_1 = 5000 \times \left( 1 - e^{-375/100\,000 \times 5} \right) = 92.877, \\
m_2 = 5000 \times \left( 1 - e^{-300/100\,000 \times 5} \right) = 74.44
\]
and \( m = m_1 + m_2 = 167.32 \). Also \( T_1 = T_2 = 5 \times 5000 = 25,000 \).

Therefore,
\[
\pi_0 = \frac{T_1}{T_1 + T_2} = 0.5, \\
\pi_A = \frac{25000 \times \frac{375}{300}}{25000 \times \frac{375}{300} + 25000} = 0.556.
\]

We wish to test \( H_0 : \rho = 1 \) vs. \( H_A : \rho \neq 1 \).

This is equivalent to testing \( H_0 : \pi = \pi_0 \) vs. \( \pi \neq \pi_0 \).

To determine the power, we use \( \pi_A \). Here
\[
V_0 = 0.5 \left( 1 - 0.5 \right) = 0.25 \quad \text{and} \quad V_A = 0.556 \left( 1 - 0.556 \right) = 0.247.
\]

Therefore,
\[
Z_1 := \frac{(\pi_A - \pi_0) \sqrt{m} - z_{1-\alpha/2} \sqrt{V_0}}{\sqrt{V_A}} \\
= \frac{(0.556 - 0.5) \sqrt{167.32} - 1.96 \sqrt{0.25}}{\sqrt{0.247}} \\
= -0.526
\]
and
\[ z_2 := \frac{(0.5 - 0.556) \sqrt{167.32} - 1.96 \sqrt{0.25}}{\sqrt{0.247}} = -3.418. \]
Thus, we obtain
\[ P(Z < z_1) + P(Z < z_2) = 0.3. \]
We will not reject a wrong null hypothesis in about 30% of the cases. Not a very good power.

Here is the code for a function that computes two-sample power for the Poisson:

```r
Poisson.power <- function(t, n, l, alpha = 0.05){
  q <- qnorm(1 - alpha / 2)
  T <- t * n
  p0 <- T[1] / sum(T) ; v0 <- p0 * (1 - p0)
  va <- pa * (1 - pa)
  m <- sum(n * (1 - exp(-l * t)))
  A <- ((pa - p0) * sqrt(m) - q * sqrt(v0)) / sqrt(va)
  B <- ((p0 - pa) * sqrt(m) - q * sqrt(v0)) / sqrt(va)
  pnorm(A) + pnorm(B)
}
```

The function takes the following arguments (except for `alpha`, all vectors are of size 2):

- `t` Time period for each sample.
- `n` Size of each sample.
- `l` $\lambda_1$ and $\lambda_2$.
- `alpha` Significance level $\alpha$ (default value = 0.05).

The function returns the two-sided power ($1 - \beta$) for a given $\alpha$. Let us follow the code for the function. In line 2 we obtain the quantile for the appropriate value of $\alpha$. In line 3, we obtain the values of subject-time for each of the sample. In our example, we have mutation-years. When the “rate” is not with respect to time, the latter represents the number of repetitions of counts for each subject. We then compute the rate ratio in line 4. In lines 5 and 6 we calculate the probability under the null and the alternative hypotheses, respectively. The variances of each sample are calculated in lines 5 and 7. The expected number of incidences (mutations in our example) are calculated in line 8. Lines 9 and 10 calculate the quantiles given in equations (13.5) and (13.6). We need both quantiles because `Poisson.power()` returns a two-sided power. Line 11 returns the power. In Exercise 13.5 you are asked to generalize the function for one-sided power (greater than and less than). 

\( \square \)
Let us discuss the sample size \( m \) that will give us a desired power. Rearranging (13.5) and (13.6) for a two-sided test, we obtain

\[
m = \left( \frac{z_{1-\alpha/2} \sqrt{V_0} + z_{1-\beta} \sqrt{V_A}}{|\pi_0 - \pi_A|} \right)^2 \tag{13.7}
\]

where \( m \) is the expected number of events in both populations. Let \( k := n_2/n_1 \). Then if we specify \( k \), we get the necessary sample sizes for each population from

\[
n_1 = \frac{m}{k + 1 - e^{-\lambda_1 t_1'} - ke^{-\lambda_2 t_2'}}, \tag{13.8}
\]

\[
n_2 = kn_1.
\]

**Example 13.6.** Continuing with Example 13.5, we ask: How many subjects do we need to follow for 5 years to obtain 80% power at significance of 0.05 for a two-tailed test and equal numbers from both populations?

Using (13.7) we write

\[
m = \left( \frac{1.96 \sqrt{0.5 (1 - 0.5) + 0.84 \sqrt{0.556 (1 - 0.556)}}}{|0.5 - 0.556|} \right)^2 = 633.40.
\]

So we need to choose \( n_1 \) and \( n_2 \) such that we anticipate 634 events to occur. From (13.8),

\[
n_1 = n_2 = \frac{634}{2 - e^{-375/100000} - e^{-300/100000} \times 5} = 18928.05.
\]

We therefore need to follow 18,929 subjects from each population for 5 years.

Here is a function that computes Poisson sample size for two samples:

```r
def Poisson.sample.size <- function(t, n, e, rho = (e[1] / n[1]) / (e[2] / n[2]), alpha = 0.05, power = 0.8, k = 1)
{
  q <- qnorm(1 - alpha / 2)
  p <- qnorm(power)
  p0 <- t[1] / sum(t); v0 <- p0 * (1 - p0)
  pa <- t[1] * rho / (t[1] * rho + t[2])
  va <- pa * (1 - pa)
  m <- (q * sqrt(v0) + p * sqrt(va)) / (abs(p0 - pa))
  m <- ceiling(m * m)
  d <- k + 1 - exp(-e[1] / n[1] * t[1]) -
  n1 <- m / d; n2 <- k * n1
  ceiling(c(n1, n2))
}
```

The function computes the sizes of two samples from Poisson populations that are necessary to achieve a given power for a given significance level and for a given ratio of the sample sizes. The function takes the following arguments:
t Time period for each sample.
\( n \) Size of each sample from past data.
\( e \) Incidence count for each sample from past data.
\( \rho \) The desired ratio of \( \lambda_1 \) to \( \lambda_2 \). If not provided, the ratio is computed from \( e \) and \( n \).
\( \alpha \) The desired significance level (default value = 0.05).
\( \text{power} \) The desired power (default value = 0.8).
\( k \) The desired ratio \( n_1 / n_2 \).

In lines 5 and 6 we compute the quantiles for \( \alpha \) and \( 1 - \beta \). In lines 7 to 9 we compute \( \pi_0 \) and \( \pi_A \) under the null and alternative hypotheses and their variances. The required number of incidences is computed in line 10 (see equation 13.7). To obtain \( n_1 \), we first compute the denominator in equation (13.8). In line 14 we compute the necessary \( n_1 \) and \( n_2 \).

13.4 Assignments

**Exercise 13.1.** Download the file `walleye.rda` from the book’s site. It contains the following list of walleye weights from two lakes:

```r
$sample.1
[1]  0.86  1.38  1.43  1.38  1.58  0.62  1.74  2.04  1.37  1.72  2.62
[12] 1.52  2.13  1.27  0.95  1.54  2.30  1.40  1.31  1.85  1.19  1.96
[23] 1.17  1.00  1.25  1.06  2.65  1.28  1.38  0.49  1.54  1.95  1.78
[34] 0.54  1.64  1.85  1.13  1.60  0.40  1.32
$sample.2
[1]  0.81  2.46  2.05  1.11  1.31  0.97  1.04  1.61  2.09  1.63  1.48
[12] 1.69  1.78  1.89  2.03  1.27  2.34  1.90  2.18  1.59  1.84  1.95
[23] 1.67  1.66  1.78  2.34  1.50  2.02  1.04  1.83  1.14  0.83  1.69
[34] 1.68  2.15  2.40  1.56  1.73  0.65  1.76  2.26  1.23  2.62  1.27
[45]  2.83
```

1. Create power profiles for the difference between the weights under the assumption that \( \mu_1 - \mu_2 = 0 \) for \( \mu_A < \mu_0, \mu_A > \mu_0 \) and \( \mu_A \neq \mu_0 \). Set the range of \( \mu_A \) from \(-1\) to \(1\).
2. What is the power of distinguishing between weights of the two samples at \( \alpha = 0.05 \) and a minimum detectable difference of 0.2 kg for \( \mu_A < \mu_0, \mu_A > \mu_0 \) and \( \mu_A \neq \mu_0 \)?

Use Example 13.1 as a guide.

**Exercise 13.2.** Continuing with the `walleye.rda` (Exercise 13.1), assume that the samples’ variances approximately equal the population variances. Set \( \alpha = 0.05 \), power between 0.6 and 0.9 and detectable difference between \(-1\) and \(1\) kg. With these:

1. Draw and interpret a figure for these data similar Figure 13.2.
2. What would be the sample size necessary to detect a difference of 0.2 kg with power = 0.9?
Exercise 13.3. Two separate populations of deer were chosen for a study of the effect of reducing winter mortality due to supplemental feeding. The first population included 38 deer and the second 42. Habitats in the two areas where the populations reside were comparable and so was the weather. The averages of the population weight at the beginning of the winter were not different. The first population received supplemental feeding, the second did not. By the end of the winter, 9 and 12 deer died from starvation in the first and second populations, respectively.

1. Was the feeding effective in reducing winter mortality?
2. What is the probability that we accept the hypothesis that the supplemental feeding was not effective in reducing mortality while in fact it was?
3. What should have been the number of deer in the winter-fed population that survived for a power of 0.8 (with $\alpha = 0.05$)?

Exercise 13.4. Continuing with Exercise 13.3, determine the population sizes that are necessary to establish a difference of 0.1 in the winter mortality between the fed and unfed deer populations. Use $\alpha = 0.05$ and $1 - \beta = 0.8$. Assign the same fraction of the total number of deer to the fed and unfed populations.

Exercise 13.5. Write a function that returns the one-sided (less than or greater than) or two-sided power of a test of the difference between $\lambda_1$ and $\lambda_2$. Use the code for `Poisson.power()` as a guide.