Statistics II : Introduction to Inference

Modules 8,9,11: Some Important Hypothesis Testing and Associated Confidence Intervals

1 Testing of binomial proportions

Example 1. *A pharmaceutical company develops a new drug intended to cure certain disease. They want to test whether the cure rate (proportion of patients cured) is better than the current standard drug, which has a known cure rate of 60%. They conduct a clinical trial where 30 patients are given the new drug. Among those, 22 are cured. Is the cure rate of the new drug significantly higher than 60%?*

(i) **Setup:** $X_1, \ldots, X_n \overset{iid}{\sim} \texttt{Bernoulli}(p)$. Consider the problem of testing $H_0 : p = p_0$ against $H_{1,1} : p > p_0$ (*or*, $H_{1,2} : p < p_0$, *or*, $H_{1,3} : p \neq p_0$). The UMP test under $H_{1,1}$ (*or*, $H_{1,2}$) at level α is

$$\phi(\boldsymbol{X}) = \begin{cases} 1 & \text{if } n\bar{X}_n > c_\alpha \text{ (or, } n\bar{X}_n < k_\alpha\text{)}, \\ \gamma & \text{if } n\bar{X}_n = c_\alpha, \\ 0 & \text{otherwise.} \end{cases}$$

Under H_0,

$$n\bar{X}_n = \sum_{i=1}^{n} X_i \sim \texttt{Binomial}(n, p_0).$$

Find the critical value c_α (or, k_α) and γ from the equation $E_{p_0}\left(\phi(\boldsymbol{X})\right) = \alpha$.

Further, under the alternative $H_{1,3}$, the above test can be generalized to

$$\phi'(\boldsymbol{X}) = \begin{cases} 1 & \text{if } n\bar{X}_n > c'_\alpha \text{ or, } n\bar{X}_n < k'_\alpha, \\ \gamma_1 & \text{if } n\bar{X}_n = c'_\alpha, \\ \gamma_2 & \text{if } n\bar{X}_n = k'_\alpha, \\ 0 & \text{otherwise.} \end{cases}$$

The critical values c'_α, k'_α and (γ_1, γ_2) should satisfy $E_{p_0}\left(\phi'(\boldsymbol{X})\right) = \alpha$. However, note that it is not possible to obtain four unknown quantities from one equation. Therefore, one imposes additional restrictions to get unique solutions. One possible restriction can be $\gamma_1 = \gamma_2$ and $c'_\alpha = n - k'_\alpha$ (symmetric critical region).

(ii) **Large sample test.** By CLT,

$$\frac{n\bar{X}_n - np_0}{\sqrt{np_0(1-p_0)}} \xrightarrow[\textit{under } H_0]{D} Z,$$

where $Z \sim N(0, 1)$.

Using this result, one can construct a large sample test for testing H_0 against the possible alternatives as follows:

$$\phi_{LS}(\boldsymbol{X}) = \begin{cases} 1 \quad \text{if} & n\bar{X}_n > c_{\alpha,LS} \text{ (against } H_{1,1}\text{)}, \\ & n\bar{X}_n < k_{\alpha,LS} \text{ (against } H_{1,2}\text{)}, \\ & n\bar{X}_n > c'_{\alpha,LS} \text{ or } n\bar{X}_n < k'_{\alpha,LS} \text{ (against } H_{1,3}\text{)}, \\ 0 \quad \text{otherwise.} & \end{cases}$$

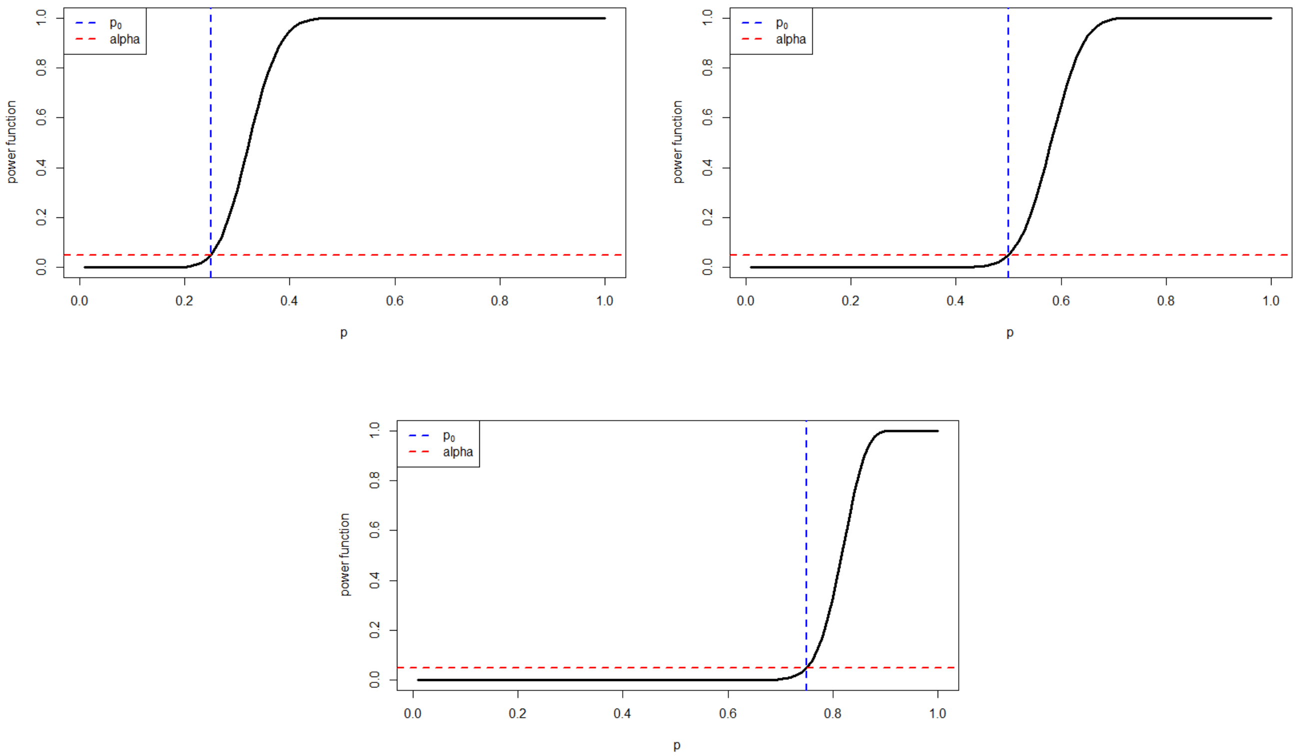

Fig. 1: Power curve for the exact test ϕ against $H_{1,1} : p > p_0$ with $p_0 = \dfrac{1}{4}$(left), $p_0 = \dfrac{1}{2}$(middle) and $p_0 = \dfrac{3}{4}$(right).

The critical points are now determined by solving the equation $E_{p_0}\left(\phi_{LS}(\boldsymbol{X})\right) = \alpha$, under the large sample distribution of $n\bar{X}_n$. For example, $c_{\alpha,LS}$ can be obtained to satisfy

$$E_{p_0}\left(\phi_{LS}(\boldsymbol{X})\right) = P_{p_0}\left(n\bar{X}_n > c_{\alpha,LS}\right) = P_{p_0}\left(\frac{n\bar{X}_n - np_0}{\sqrt{np_0(1-p_0)}} > \frac{c_{\alpha,LS} - np_0}{\sqrt{np_0(1-p_0)}}\right) = \alpha,$$

which provides $c_{\alpha,LS} = np_0 + \tau_\alpha\sqrt{np_0(1-p_0)}$. Here τ_α is the upper-α point of $N(0,1)$ distribution, i.e., τ_α is such that $P(X \geq \tau_\alpha \mid X \sim N(0,1)) = \alpha$.

Similarly, $k_{\alpha,LS}$ can be obtained to satisfy

$$E_{p_0}\left(\phi_{LS}(\boldsymbol{X})\right) = P_{p_0}\left(n\bar{X}_n < k_{\alpha,LS}\right) = P_{p_0}\left(\frac{n\bar{X}_n - np_0}{\sqrt{np_0(1-p_0)}} < \frac{k_{\alpha,LS} - np_0}{\sqrt{np_0(1-p_0)}}\right) = \alpha,$$

which provides $k_{\alpha,LS} = np_0 - \tau_\alpha\sqrt{np_0(1-p_0)}$, as $\tau_{1-\alpha} = -\tau_\alpha$.

Finally, $(c'_{\alpha,LS}, k'_{\alpha,LS})$ are such that they satisfy

$$E_{p_0}\left(\phi_{LS}(\boldsymbol{X})\right) = P_{p_0}\left(n\bar{X}_n < k'_{\alpha,LS}\right) + P_{p_0}\left(n\bar{X}_n > c'_{\alpha,LS}\right) = \alpha.$$

As it is not possible to uniquely solve for $(c'_{\alpha,LS}, k'_{\alpha,LS})$ from the above equation, we put the additional restriction that

$$P_{p_0}\left(n\bar{X}_n < k'_{\alpha,LS}\right) = \alpha/2, \qquad \text{and} \qquad P_{p_0}\left(n\bar{X}_n > c'_{\alpha,LS}\right) = \alpha/2,$$

and obtain the solutions

$$k'_{\alpha,LS} = np_0 - \tau_{\alpha/2}\sqrt{np_0(1-p_0)}, \qquad \text{and} \qquad c'_{\alpha,LS} = np_0 + \tau_{\alpha/2}\sqrt{np_0(1-p_0)}.$$

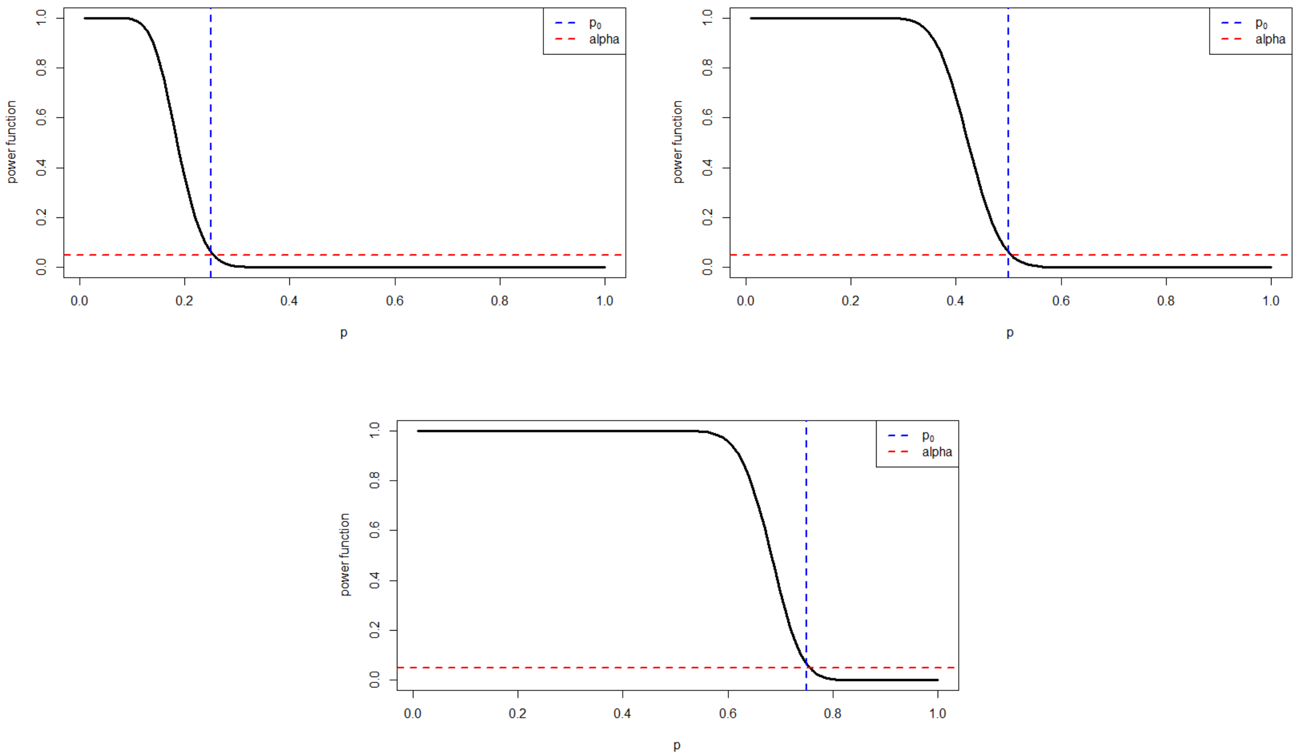

Fig. 2: Power curve for the exact test ϕ against $H_{1,2} : p < p_0$ with $p_0 = \dfrac{1}{4}$(left), $p_0 = \dfrac{1}{2}$(middle) and $p_0 = \dfrac{3}{4}$(right).

(iii) **Power function.** The power curve for the exact test ϕ (under the alternative $H_{1,1}$) for three choices of $p_0 \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$ is given in Figure 1.

The power curve for the exact test ϕ (under the alternative $H_{1,2}$) for three choices of $p_0 \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$ is given in Figure 2.

(iv) **Large sample confidence interval associated with the test.** The large sample test ϕ_{LS} under the alternative $H_{1,3}$ leads to a possible (large sample) confidence interval for binomial proportion p. Under $H_0 : p = p_0$,

$$\frac{n\bar{X}_n - np_0}{\sqrt{np_0(1-p_0)}} \xrightarrow[under\ \ H_0]{D} Z,$$

where $Z \sim N(0,1)$. Note that $\bar{X}_n \xrightarrow{p} p_0$ (by WLLN), and so by continuous mapping

$$\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{p_0(1-p_0)}} \xrightarrow{p} 1.$$

Therefore, by Slutsky's theorem

$$\frac{\sqrt{n}(\bar{X}_n - p_0)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow[under\ \ H_0]{D} Z,$$

where $Z \sim N(0,1)$. Based on this result one can obtain that

$$P_{p_0}\left(-\tau_{\alpha/2} \leq \frac{\sqrt{n}(\bar{X}_n - p_0)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \leq \tau_{\alpha/2}\right) \approx 1 - \alpha,$$

which is equivalent to saying

$$\begin{aligned} & P_p\left(-\tau_{\alpha/2} \leq \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \leq \tau_{\alpha/2}\right) && \approx 1-\alpha, \\ \Leftrightarrow \quad & P_p\left(\bar{X}_n - \tau_{\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} \leq p \leq \bar{X}_n + \tau_{\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right) && \approx 1-\alpha. \end{aligned}$$

for sufficiently large n.

Therefore, $\left[\bar{X}_n - \tau_{\alpha/2}\sqrt{\bar{X}_n(1-\bar{X}_n)/n}, \bar{X}_n + \tau_{\alpha/2}\sqrt{\bar{X}_n(1-\bar{X}_n)/n}\right]$ is a confidence interval for p with confidence coefficient $1-\alpha$.

2 Testing of normal mean (Student t-test)

Example 2. *A researcher wants to find out if university students are getting less than the recommended 8 hours of sleep per night. She randomly samples 25 students and records their sleep hours. The sample mean turns out to be 7.5 hours, and the sample standard deviation is 1.2 hours. Is there enough evidence to conclude that the students sleep less than 8 hours on average?*

(i) **Setup:** $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, where σ^2 is unknown. We want to test $H_0 : \mu = \mu_0$ against $H_{1,1} : \mu > \mu_0$ $(or,\ H_{1,2} : \mu < \mu_0,\ or,\ H_{1,3} : \mu \neq \mu_0)$.

- We obtained UMP test against $H_{1,1}$ and $H_{1,2}$ when σ^2 is known with the test statistic $\bar{X}_n$.
- In the same spirit, let us construct the following test for H_0 against $H_{1,1}$ $(or,\ H_{1,2})$ as

$$\phi(\boldsymbol{X}) = \begin{cases} 1 & \text{if } \bar{X}_n \geq c_\alpha, \\ 0 & \text{otherwise.} \end{cases}$$

However, the problem arises in finding the critical value as

$$P_{H_0}(\bar{X}_n \geq c_\alpha) = \alpha \Leftrightarrow 1 - \Phi\left(\frac{c_\alpha - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha,$$

which can not be solved when σ is unknown.

- Therefore, we need a test statistic which is a function of $\bar{X}_n$, but it's distribution will be completely known under H_0.
- Define

$$T_n = \frac{\bar{X}_n - \mu_0}{\hat{\sigma}_n/\sqrt{n}}, \tag{1}$$

where $\hat{\sigma}_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

As $Z_n = \sqrt{n}(\bar{X}_n - \mu_0)/\sigma \overset{H_0}{\sim} N(0,1)$, $W_n = (n-1)\hat{\sigma}_n^2/\sigma^2 \sim \chi^2_{n-1}$ and Z_n is independent of W_n, we have

$$\frac{Z_n}{\sqrt{W_n/(n-1)}} = T_n \overset{H_0}{\sim} t_{n-1} \quad \text{distribution.}$$

- **Test procedure.** Based on the proposed t-statistic the test function for testing H_0 against the possible alternatives is as follows:

$$\phi(\boldsymbol{X}) = \begin{cases} 1 \quad \text{if} & T_n > c_\alpha \ \text{(against } H_{1,1}),\\ & T_n < k_\alpha \ \text{(against } H_{1,2}),\\ & T_n > c'_\alpha \ \text{ or } \ T_n < k'_\alpha \ \text{(against } H_{1,3}),\\ 0 \quad \text{otherwise.} & \end{cases}$$

The critical points are now determined by solving the equation $E_{\mu_0}(\phi(\boldsymbol{X})) = \alpha$, under the distribution of T_n under H_0. In particular, $c_\alpha = t_{\alpha,n-1}$ and $k_\alpha = t_{1-\alpha,n-1} = -t_{\alpha,n-1}$, where $t_{\alpha,n-1}$ is the upper-α point of t_{n-1} distribution, i.e., $t_{\alpha,n-1}$ is such that $P(T \geq t_{\alpha,n-1} \mid T \sim t_{n-1}) = \alpha$. Further, under the additional assumption of symmetric critical region, one obtains $c'_\alpha = t_{\alpha/2,n-1}$ and $k'_\alpha = t_{1-\alpha/2,n-1} = -t_{\alpha/2,n-1}$,
- The test statistic in (1) is called the *student t statistic*, and the corresponding test is called the **student t-test**.

(ii) **Power function.** Define $\mu - \mu_0 = \Delta$. Under H_0, $\Delta = 0$, and under the alternatives it can take any value in $\mathbb{R} \setminus \{0\}$. The power function of the test ϕ under the alternative H_{11} can be derived as follows:

$$\begin{aligned} \beta_\phi(\mu) &= P_\mu\left(T_n > t_{\alpha,n-1} \ \mid \ T_n + \frac{\sqrt{n}\Delta}{\hat{\sigma}_n} \sim t_{n-1}\right) \\ &= P_{\sigma^2}\left(T_n + \frac{\sqrt{n}\Delta}{\hat{\sigma}_n} > t_{\alpha,n-1} + \frac{\sqrt{n}\Delta}{\hat{\sigma}_n} \ \mid \ T_n + \frac{\sqrt{n}\Delta}{\hat{\sigma}_n} \sim t_{n-1}\right) \\ &= 1 - G\left(t_{\alpha,n-1} + \frac{\sqrt{n}\Delta}{\hat{\sigma}_n}\right), \end{aligned}$$

where G is the CDF of t_{n-1} distribution.

Similarly, one can derive the power function of ϕ under the alternatives $H_{1,2}$ and $H_{1,3}$. The power curve of ϕ under the alternative $H_{1,3}$ for $\mu_0 = 10$ is given in Figure 3.

(iii) **Large sample test.** Let $T_n \sim t_n$ distribution. Then as $n \to \infty$, $T_n \xrightarrow{d} Z$ where $Z \sim N(0,1)$.

To see this, let $X, X_1, \ldots, X_n \overset{iid}{\sim} N(0,1)$. Then

$$T_n \overset{D}{=} \frac{X}{\sqrt{n^{-1}(X_1^2 + \ldots + X_n^2)}} \sim t_n,$$

where $U \overset{D}{=} V$ implies U, V have same distribution (U and V are not necessarily same variables). By WLLN and continuous mapping $\sqrt{n^{-1}(X_1^2 + \ldots + X_n^2)} \xrightarrow{p} 1$ (verify), and therefore by Slutsky's lemma, $T_n \xrightarrow{D} Z$ where $Z \sim N(0,1)$.

Based on the above results a large sample version of the above t-test can be obtained as follows:

$$\phi_{LS}(\boldsymbol{X}) = \begin{cases} 1 \quad \text{if} & T_n > c_{\alpha,LS} \ \text{(against } H_{1,1}),\\ & T_n < k_{\alpha,LS} \ \text{(against } H_{1,2}),\\ & T_n > c'_{\alpha,LS} \ \text{ or } \ T_n < k_\alpha^{\prime,LS} \ \text{(against } H_{1,3}),\\ 0 \quad \text{otherwise.} & \end{cases}$$

The critical points are now determined by solving the equation $E_{\mu_0}(\phi_{LS}(\boldsymbol{X})) = \alpha$, under the large sample distribution of T_n under H_0. In particular, $c_{\alpha,LS} = \tau_\alpha$, $k_{\alpha,LS} = \tau_{1-\alpha} = -\tau_\alpha$, and $c'_{\alpha,LS} = \tau_{\alpha/2}$, $k'_{\alpha,LS} = \tau_{1-\alpha/2} = -\tau_{\alpha/2}$ (the last two obtained under the additional restriction of symmetric critical region), where τ_α is the upper-α point of $N(0,1)$ distribution.

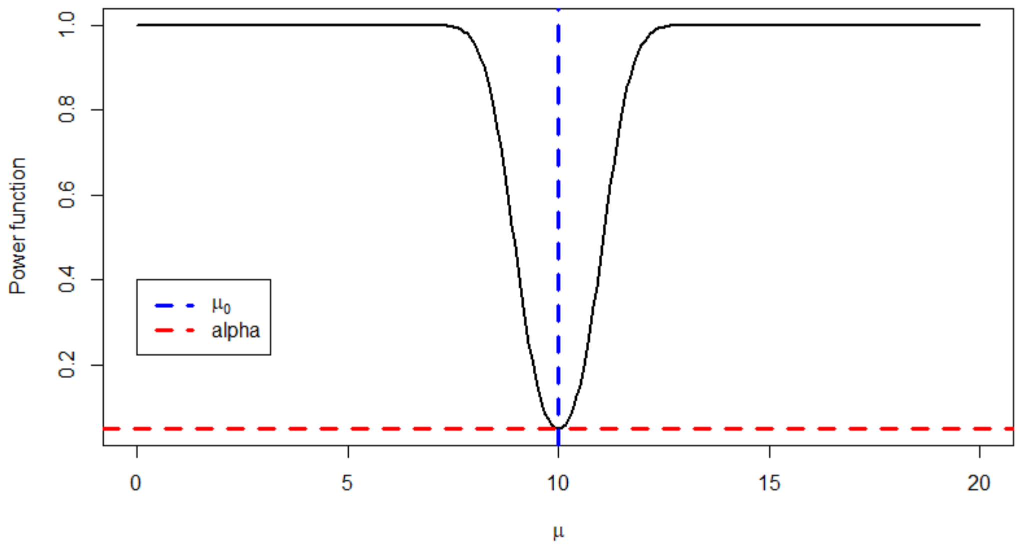

Fig. 3: Power curve for $H_{1,3} : \mu \neq \mu_0$ with $\mu_0 = 10$.

(iv) **Confidence interval.** By inverting the test statistic of the exact or large sample test of $H_0 : \mu = \mu_0$ against $H_{1,3} : \mu \neq \mu_0$, one can obtain a confidence interval for μ.

From the exact test above, we obtain

$$
\begin{aligned}
P_{\mu_0}\left(k'_\alpha \leq T_n \leq c'_\alpha\right) &= P_{\mu_0}\left(-t_{\alpha/2,n-1} \leq \frac{\bar{X}_n - \mu_0}{\hat{\sigma}_n/\sqrt{n}} \leq t_{\alpha/2,n-1}\right) = 1-\alpha, \\
&\Leftrightarrow P_{\mu}\left(-t_{\alpha/2,n-1} \leq \frac{\bar{X}_n - \mu}{\hat{\sigma}_n/\sqrt{n}} \leq t_{\alpha/2,n-1}\right) = 1-\alpha, \\
&\Leftrightarrow P_{\mu}\left(\bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} t_{\alpha/2,n-1} \leq \mu \leq \bar{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} t_{\alpha/2,n-1}\right) = 1-\alpha.
\end{aligned}
$$

Therefore, $\left[\bar{X}_n - \hat{\sigma}_n t_{\alpha/2,n-1}/\sqrt{n}, \bar{X}_n + \hat{\sigma}_n/\sqrt{n} t_{\alpha/2,n-1}\right]$ is a confidence interval for μ with confidence coefficient $1-\alpha$. Similarly, one can obtain a large sample $(1-\alpha)$ confidence interval for μ by replacing $t_{\alpha/2,n-1}$ in the above confidence interval by $\tau_{\alpha/2}$.

2.1 Paired-t test

Example 3. *The marks obtained by 10 students in a mock test and an actual test are given below. Based on this data, is there sufficient evidence to conclude that students score significantly lower in the mock test compared to the actual test?*

Student ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Marks in Mock Test	45	38	40	36	41	39	35	37	42	44
Marks in Actual Test	50	42	47	38	44	43	40	42	46	47

(i) **Setup.** Let $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n \overset{iid}{\sim} \mathcal{N}_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \sigma_{XY})$ (*i.e.*, Bivariate normal distribution) and $\boldsymbol{Z}_i =$

$\begin{pmatrix} X_i \\ Y_i \end{pmatrix}$, for $i = 1, 2, \ldots, n$, which implies

$$X_i \overset{iid}{\sim} \mathcal{N}(\mu_X, \sigma_X^2), \quad Y_i \overset{iid}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2) \quad \text{and} \quad Cov(X_i, Y_i) = \sigma_{XY}.$$

To test

$$H_0 : \mu_X = \mu_Y \quad (i.e.,\ \mu_X - \mu_Y = 0)$$
$$\text{against}$$
$$H_{1,1} : \mu_X - \mu_Y > 0 \quad (or,\ H_{1,2} : \mu_X - \mu_Y < 0,\ or,\ H_{1,3} : \mu_X - \mu_Y \neq 0).$$

(ii) **Test statistic.** Consider the difference of two variables $W_i = X_i - Y_i$. By properties of bivariate normal distribution $W_i \sim \mathcal{N}(\mu_W, \sigma_W^2)$, where $\mu_W = \mu_X - \mu_Y$ and $\sigma_W^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$.

Therefore, the testing problem can be regarded as a special case of the testing of normal mean with variance unknown, i.e., $H_0 : \mu_W = 0$ against possible alternatives.

The related test statistic is

$$T_n = \frac{\bar{X}_n - \bar{Y}_n}{\hat{\sigma}_{W,n}/\sqrt{n}}, \tag{2}$$

where $\hat{\sigma}_{W,n}^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - Y_i - \bar{X}_n + \bar{Y}_n)^2$.

(iii) The test statistic in (2) is called the *paired t-test statistic*, and the corresponding test is called the **paired t-test**. Further, the assumption of same variances for the two normal populations, i.e., $\text{var}(X_1) = \text{var}(Y_1) = \sigma^2$ is called the *homoscedasticity* assumption.

3 Testing of normal variance

Example 4. *A factory produces ball bearings with a target diameter of* 10 *mm. The manufacturing process is designed to keep the variation (standard deviation) in the diameters small, say with a target variance of* 0.01 *mm*^{2}*. The quality control (QC) team suspects that a recent change in materials or machine calibration may have increased the variability of the product. They collect a random sample of* 25 *ball bearings and measure their diameters. Is there sufficient evidence to conclude that the variance in ball bearing diameters has increased from the target value of* 0.01 *mm*^{2}*?*

(i) **Setup:** $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, where μ is unknown. We want to test $H_0 : \sigma^2 = \sigma_0^2$ against $H_{1,1} : \sigma^2 > \sigma_0^2$ $(or,\ H_{1,2} : \sigma^2 < \sigma_0^2,\ or,\ H_{1,3} : \sigma^2 \neq \sigma_0^2)$.

- We obtained UMP test against $H_{1,1}$ and $H_{1,2}$ with the test statistic $nS_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2$ (verify starting from Neyman Pearson's lemma, and then by generalizing).
- In the same spirit, we can propose the following test for testing H_0 against H_{13}.

$$\phi(\boldsymbol{X}) = \begin{cases} 1 & \text{if} \quad nS_n^2 \geq c_\alpha, \quad or, \quad nS_n^2 \leq k_\alpha \\ 0 & \text{otherwise.} \end{cases}$$

The critical points are now determined by solving the equation $E_{\sigma_0^2}(\phi(\boldsymbol{X})) = \alpha$, under the distribution of T_n under H_0. As

$$\frac{nS_n^2}{\sigma_0^2} \overset{H_0}{\sim} \chi_{n-1}^2,$$

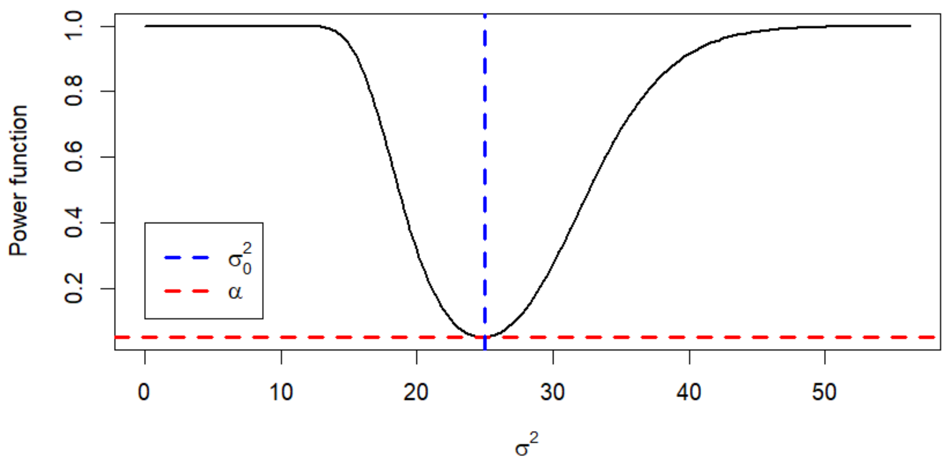

Fig. 4: The power curve of the test function ϕ for testing $H_0 : \sigma^2 = 25$ against $H_{13} : \sigma^2 \neq 25$.

the above equation provides

$$P_{\sigma_0^2}\left(\frac{nS_n^2}{\sigma_0^2} \geq c_\alpha\right) + P_{\sigma_0^2}\left(\frac{nS_n^2}{\sigma_0^2} \leq k_\alpha\right) = \alpha.$$

As two unknowns (c_α, k_α) can not be solved from one equation, we additionally impose the restriction of symmetry as follows:

$$P_{\sigma_0^2}\left(\frac{nS_n^2}{\sigma_0^2} \geq c_\alpha\right) = \frac{\alpha}{2}, \qquad P_{\sigma_0^2}\left(\frac{nS_n^2}{\sigma_0^2} \leq k_\alpha\right) = \frac{\alpha}{2},$$

and obtain the solutions $c_\alpha = \sigma_0^2 \chi^2_{\alpha/2,n-1}$, $k_\alpha = \sigma_0^2 \chi^2_{1-\alpha/2,n-1}$, where $\chi^2_{\alpha,n-1}$ is the upper-α point of χ^2_{n-1} distribution, i.e., $\chi^2_{\alpha,n-1}$ is such that $P(T \geq \chi^2_{\alpha,n-1} \mid T \sim \chi^2_{n-1}) = \alpha$.

(ii) **Power function.** The power curve of the above test can be derived as

$$\begin{aligned} \beta_\phi(\sigma^2) &= P_{\sigma^2}\left(\frac{nS_n^2}{\sigma^2} \geq \frac{\sigma_0^2}{\sigma^2}\chi^2_{\alpha/2,n-1}\right) + P_{\sigma^2}\left(\frac{nS_n^2}{\sigma^2} \leq \frac{\sigma_0^2}{\sigma^2}\chi^2_{1-\alpha/2,n-1}\right) \\ &= 1 - G\left(\frac{\sigma_0^2}{\sigma^2}\chi^2_{\alpha/2,n-1}\right) + G\left(\frac{\sigma_0^2}{\sigma^2}\chi^2_{1-\alpha/2,n-1}\right), \end{aligned}$$

where G is the CDF of χ^2_{n-1} distribution. The power curve of the test function for testing $H_0 : \sigma^2 = 25$ against $H_{13} : \sigma^2 \neq 25$ is given in Figure 4.

(iii) **Confidence interval.** By inverting the test statistic of the test of $H_0 : \sigma^2 = \sigma_0^2$ against $H_{1,3} : \sigma^2 \neq \sigma_0^2$, one can obtain a confidence interval for σ^2.

From the exact test above, we obtain

$$\begin{aligned} P_{\sigma_0^2}(k_\alpha \leq nS_n \leq c_\alpha) &= P_{\sigma_0^2}\left(\sigma_0^2\chi^2_{1-\alpha/2,n-1} \leq nS_n^2 \leq \sigma_0^2\chi^2_{1-\alpha/2,n-1}\right) = 1-\alpha, \\ &\Leftrightarrow P_{\sigma^2}\left(\sigma^2\chi^2_{1-\alpha/2,n-1} \leq nS_n^2 \leq \sigma^2\chi^2_{1-\alpha/2,n-1}\right) = 1-\alpha, \\ &\Leftrightarrow P_{\sigma^2}\left(\frac{nSn^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{nSn^2}{\chi^2_{1-\alpha/2,n-1}}\right) = 1-\alpha. \end{aligned}$$

Therefore, $\left[\dfrac{nSn^2}{\chi^2_{\alpha/2,n-1}}, \dfrac{nSn^2}{\chi^2_{1-\alpha/2,n-1}}\right]$ is a confidence interval for σ^2 with confidence coefficient $1-\alpha$.

4 Two-sample normal mean test

Example 5. *A school wants to know if a new teaching method improves student test scores compared to the traditional method.*

- *Group 1: 30 students taught with the new method. Sample mean score = 78, sample standard deviation = 10.*
- *Group 2: 28 students taught with the traditional method. Sample mean score = 74, sample standard deviation = 11.*

Is there a significant difference in mean test scores between the two methods?

(i) **Setup:** $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ and $Y_1, \ldots, Y_m \overset{iid}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$. Further, $\{X_1, \ldots, X_n, Y_1, \ldots, Y_m\}$ are mutually independent, and σ^2 is unknown.

We want to test $H_0 : \mu_X = \mu_Y$ against $H_{1,1} : \mu_X > \mu_Y$ $(or,\ H_{1,2} : \mu_X < \mu_Y,\ or,\ H_{1,3} : \mu_X \neq \mu_Y)$.

(ii) **Test statistic**:

- As before, we need a test statistic based on $(\bar{X}_n - \bar{Y}_m)$. [However, it does not make sense to consider $(X_i - Y_i)$ as the samples are not paired.]
- Further, we need a test statistic whose distribution, under H_0, is completely known.
- Observe that

$$\bar{X}_n \sim N\left(\mu_X, \frac{\sigma^2}{n}\right), \qquad \text{and} \qquad \bar{Y}_m \sim N\left(\mu_Y, \frac{\sigma^2}{m}\right), \quad \text{independently, which}$$

implies $\quad \bar{X}_n - \bar{Y}_m \sim N\left(\mu_X - \mu_Y, \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right)$, i.e., $\dfrac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim N(0, 1)$.

- Also,

$$\frac{nS_X^2}{\sigma^2} \sim \chi^2_{n-1}, \quad \text{and} \quad \frac{mS_Y^2}{\sigma^2} \sim \chi^2_{m-1}, \quad \text{independently, implying,} \quad \frac{nS_X^2 + mS_Y^2}{\sigma^2} \sim \chi^2_{n+m-2},$$

where S_X^2 and S_Y^2 are sample variances of $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_m\}$, respectively, i.e., $nS_X^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2$, $mS_Y^2 = \sum_{j=1}^m (Y_j - \bar{Y}_m)^2$

- Finally, note that all the four quantities $\{\bar{X}_n, \bar{Y}_m, S_X^2, S_Y^2\}$ are mutually independent (WHY ?), which implies that $\bar{X}_n - \bar{Y}_m$ is independent of $nS_X^2 + mS_Y^2$.
- Combining the last three points, we get

$$T_{n,m} = \frac{\bar{X}_n - \bar{Y}_m}{\hat{\sigma}_{n,m}\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \Bigg/ \sqrt{\frac{nS_X^2 + mS_Y^2}{\sigma^2(n+m-2)}} \sim t_{(n+m-2)}, \tag{3}$$

where $\hat{\sigma}^2_{n,m} = (n+m-2)^{-1}(nS_X^2 + mS_Y^2)$.

Therefore, under H_0,

$$\frac{\bar{X}_n - \bar{Y}_M}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \Bigg/ \sqrt{\frac{nS_X^2 + mS_Y^2}{\sigma^2(n+m-2)}} \overset{H_0}{\sim} t_{m+n-2}.$$

(iii) Derive the complete test function, the power function and corresponding confidence interval based on the test based on the above test statistic.

(iv) The test statistic in (3) is called the *two-sample student t-test statistic*, and the corresponding test is called the **two-sample student t-test**.

(v) Finally when $\min\{m, n\} \to \infty$, an asymptotic version of the above test can be formed, where the critical values are obtained from the normal distribution.

5 Testing equality of two normal variances

Example 6. *A factory uses two different machines (Machine A and Machine B) to produce metal rods. The management wants to check if the consistency (variation in rod length) is the same for both machines.*

- *From Machine A, they take a sample of 20 rods and find a sample variance of 0.04 cm^2.*
- *From Machine B, they take a sample of 25 rods and find a sample variance of 0.06 cm^2.*

Is there a significant difference between the variances of the two machines?

(i) **Setup:** $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, \ldots, Y_m \overset{iid}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, and $\{X_1, \ldots, X_n, Y_1, \ldots, Y_m\}$ are mutually independent, where μ_X, μ_Y, σ_X^2 and σ_Y^2 are unknown.

We want to test

$$H_0 : \sigma_X^2 = \sigma_Y^2 \Leftrightarrow \frac{\sigma_X^2}{\sigma_Y^2} = 1 \text{ against } H_{1,1} : \frac{\sigma_X^2}{\sigma_Y^2} > 1 \ (or,\ H_{1,2} : \frac{\sigma_X^2}{\sigma_Y^2} < 1,\ \ or,\ H_{1,3} : \frac{\sigma_X^2}{\sigma_Y^2} \neq 1).$$

(ii) **Test statistic.**

- As the test concerns ratio of the population variances, it is natural to consider a test statistic based on the ratio of the sample variances $\dfrac{S_X^2}{S_Y^2}$.
- Recall that

$$\frac{nS_X^2}{\sigma_X^2} \sim \chi^2_{n-1}, \quad \text{and} \quad \frac{mS_Y^2}{\sigma_Y^2} \sim \chi^2_{m-1}, \quad \text{independently, implying,} \quad \frac{nS_X^2/(n-1)}{mS_Y^2/(m-1)} \frac{\sigma_Y^2}{\sigma_X^2} \sim \chi^2_{n+m-2}.$$

Therefore, under H_0, we have

$$T_{n,m} = \frac{nS_X^2/(n-1)}{mS_Y^2/(m-1)} \overset{H_0}{\sim} F_{n-1,m-1}.$$

(iii) Based on the above test statistic, the following test function can be derived:

$$\phi(\boldsymbol{X}) = \begin{cases} 1 \quad \text{if} & T_{n,m} > c_\alpha \ \ \text{(against } H_{1,1}), \\ & T_{n,m} < k_\alpha \ \ \text{(against } H_{1,2}), \\ & T_{n,m} > c'_\alpha \ \text{ or } \ T_n < k'_\alpha \ \ \text{(against } H_{1,3}), \\ 0 \quad \text{otherwise.} \end{cases}$$

The critical points are now determined by solving the equation $E_{H_0}(\phi(\boldsymbol{X})) = \alpha$, under the distribution of T_n under H_0. In particular, $c_\alpha = F_{\alpha,n-1,m-1}$, $k_\alpha = F_{1-\alpha,n-1,m-1}$, and $c'_\alpha = F_{\alpha/2,n-1,m-1}$, $k'_\alpha = F_{1-\alpha/2,n-1,m-1}$, where $F_{\alpha,n-1,m-1}$ is the upper-α point of $F_{n-1,m-1}$ distribution, i.e., $F_{\alpha,n-1,m-1}$ is such that $P(F \geq F_{\alpha,n-1,m-1} \mid F \sim F_{n-1,m-1}) = \alpha$.

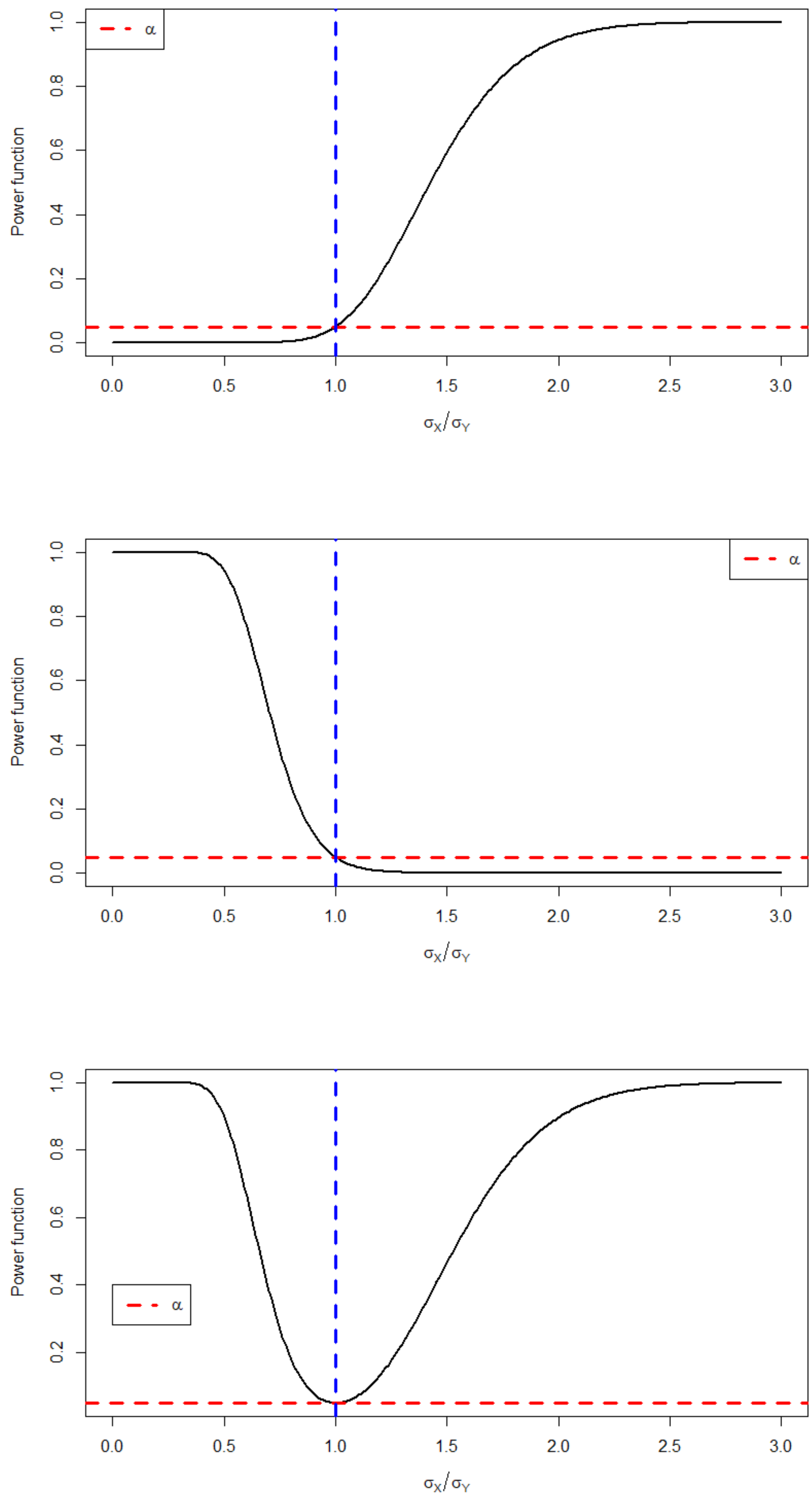

Fig. 5: Power function for the F-test for $H_0 : \sigma_X^2 = \sigma_Y^2$ against $H_{1,1} : \sigma_X^2 > \sigma_Y^2$ (top), $H_{1,1} : \sigma_X^2 < \sigma_Y^2$ (middle), and $H_{1,3} : \sigma_X^2 \neq \sigma_Y^2$ (bottom).

(iv) **Power function.** Define $\sigma_X^2/\sigma_Y^2 = \sigma^2$. Under H_0, $\sigma^2 = 1$, and other the alternative it can take any positive value. The power function of the test ϕ under the alternative H_{11} can be derived as follows:

$$\begin{aligned}\beta_\phi(\sigma^2) &= P_{\sigma^2}\left(T_{m,n} > F_{\alpha,n-1,m-1} \mid T_{m,n}/\sigma^2 \sim F_{n-1,m-1}\right)\\ &= P_{\sigma^2}\left(T_{m,n}/\sigma^2 > F_{\alpha,n-1,m-1}/\sigma^2 \mid T_{m,n}/\sigma^2 \sim F_{n-1,m-1}\right)\\ &= 1 - G\left(F_{\alpha,n-1,m-1}/\sigma^2\right),\end{aligned}$$

where G is the CDF of $F_{n-1,m-1}$ distribution.

Similarly, one can derive the power function of ϕ under the alternatives $H_{1,2}$ and $H_{1,3}$. The graph of the power function of ϕ under possible alternatives are given in Figure 5.

(v) **Confidence interval.** Based on the above test one can form a confidence interval for the ratio of variances of two normal populations. From the test we get

$$\begin{aligned}P_{\sigma_X^2/\sigma_Y^2=1}\left(k'_\alpha \le T_{m,n} \le c'_\alpha\right) &= P_{\sigma_X^2/\sigma_Y^2=1}\left(F_{1-\alpha/2,n-1,m-1} \le \frac{nS_X^2/(n-1)}{mS_Y^2/(m-1)} \le F_{\alpha/2,n-1,m-1}\right)\\ &= 1-\alpha.\end{aligned}$$

For a general σ_X^2/σ_Y^2 (not necessarily 1), we have

$$\begin{aligned}P_{\sigma_X^2/\sigma_Y^2}\left(F_{1-\alpha/2,n-1,m-1} \le \frac{nS_X^2/(n-1)}{mS_Y^2/(m-1)}\frac{\sigma_Y^2}{\sigma_X^2} \le F_{\alpha/2,n-1,m-1}\right) &= 1-\alpha\\ P_{\sigma_X^2/\sigma_Y^2}\left(\frac{nS_X^2/(n-1)}{mS_Y^2/(m-1)F_{\alpha/2,n-1,m-1}} \le \frac{\sigma_X^2}{\sigma_Y^2} \le \frac{nS_X^2/(n-1)}{mS_Y^2/(m-1)F_{1-\alpha/2,n-1,m-1}}\right) &= 1-\alpha.\end{aligned}$$

Therefore, $\left[\frac{nS_X^2/(n-1)}{mS_Y^2/(m-1)F_{\alpha/2,n-1,m-1}}, \frac{nS_X^2/(n-1)}{mS_Y^2/(m-1)F_{1-\alpha/2,n-1,m-1}}\right]$ is a confidence interval for σ_X^2/σ_Y^2 with confidence coefficient $1-\alpha$.

6 Exercises

1. Suppose that the proportion p of defective items in a large population of items is unknown, and that it is desired to test the following hypotheses:

$$H_0 : p = 0.2 \quad \text{vs} \quad H_1 : p \neq 0.2.$$

   Suppose also that a random sample of 20 items is drawn from the population. Let Y denote the number of defective items in the sample, and consider a test procedure ϕ such that the critical region contains all the outcomes for which either $Y \geq 7$ or $Y \leq 1$.

   (a) Determine the value of the power function $\beta_\phi(p)$ at the points $p = 0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1; and sketch the power function.

   (b) Determine the size of the test.

2. An auto manufacturer gives a warranty for 3 years for its new vehicles. In a random sample of 60 of its vehicles, 20 of them needed five or more major repairs within the warranty period. Estimate the 95% (large sample) confidence interval of the true proportion of vehicles from this manufacturer that need five or more major repairs during the warranty period, with confidence coefficient 0.95. Interpret the result.

3. Suppose that a random sample of 10,000 observations is taken from the normal distribution with unknown mean μ and known variance is 1, and it is desired to test the following hypotheses at the level of significance 0.05:

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0.$$

   Suppose also that the test procedure specifies rejecting H_0 when $|\bar{X}_n| \geq c$, where the constant c is chosen so that $P(|\bar{X}_n| \geq c \mid \mu = 0) = 0.05$. Find the probability that the test will reject H_0 if (a) the actual value of μ is 0.01, and (b) the actual value of μ is 0.02.

4. A random sample of size 50 from a particular brand of tea packets produced a mean weight of 15.65 ounces and standard deviation of 0.59 ounce. Assume that the weights of these brands of tea packets are normally distributed. Find a 95% confidence interval for the true mean μ.

5. Fifteen vehicles were observed at random for their speeds (in mph) on a highway with speed limit posted as 70 mph, and it was found that their average speed was 73.3 mph. Suppose that from past experience we can assume that vehicle speeds are normally distributed with $\sigma = 3.2$. Construct a 90% confidence interval for the true mean speed μ, of the vehicles on this highway. Interpret the result.

6. Studies have shown that the risk of developing coronary disease increases with the level of obesity, or accumulation of body fat. A study was conducted on the effect of exercise on losing weight. Fifty men who exercised lost an average of 11.4 lb, with a standard deviation of 4.5 lb. Construct a 95% confidence interval for the mean weight loss through exercise. Interpret the result and state any assumptions you have made.

7. Two statistics professors want to estimate average scores for an elementary statistics course that has two sections. Each professor teaches one section and each section has a large number of students. A random sample of 50 scores from each section produced the following results:

   (a) Section I: $\bar{x}_1 = 77.01, s_1 = 10.32$

   (b) Section II: $\bar{x}_2 = 72.22, s_2 = 11.02$

   Calculate 95% confidence intervals for each of these three samples.

8. Suppose that $X_1, \dots, X_m$ form a random sample from the normal distribution with unknown mean μ_1 and unknown variance σ_1^2, and that $Y_1, \dots, Y_n$ form an independent random sample from the normal distribution with unknown mean μ_2 and unknown variance σ_2^2. Suppose also that it is desired to test the following hypotheses with the usual F-test at the level of significance $\alpha = 0.05$:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 > \sigma_2^2.$$

Assuming that $m = 16$ and $n = 21$, show that the power of the test when $\sigma_1^2 = 2\sigma_2^2$ is given by $P(V \geq 1.1)$, where V is a random variable having the F-distribution with 15 and 20 degrees of freedom.

9. The scores of a random sample of 16 people who took the TOEFL (Test of English as a Foreign Language) had a mean of 540 and a standard deviation of 50. Construct a 95% confidence interval for the population mean μ of the TOEFL score, assuming that the scores are normally distributed.

10. The following data represent the rates (micrometers per hour) at which a razor cut made in the skin of anesthetized newts is closed by new cells.

$$28, 20, 21, 39, 32, 23, 18, 31, 14, 23, 18, 22, 28, 24, 33, 12, 23, 21, 25, \text{ and } 25$$

    (a) Find the 95% confidence interval for population mean rate μ for the new cells to close a razor cut made in the skin of anesthetized newts.

    (b) Find a 99% confidence interval for μ. Is the 95% CI wider or narrower than the 99% CI? Briefly explain why.

    (c) Find the 95% confidence interval for population variance σ^2.

11. A study of two kinds of machine failures shows that 58 failures of the first kind took on the average 79.7 minutes to repair with a standard deviation of 18.4 minutes, whereas 71 failures of the second kind took on average 87.3 minutes to repair with a standard deviation of 19.5 minutes. Find a 99% confidence interval for the difference between the true average amounts of time it takes to repair failures of the two kinds of machines.

12. The management of a supermarket wanted to study the spending habits of its male and female customers. A random sample of 16 male customers who shopped at this supermarket showed that they spent an average of \$55 with a standard deviation of \$12. Another random sample of 25 female customers showed that they spent \$85 with a standard deviation of \$20.50. Assuming that the amounts spent at this supermarket by all its male and female customers were approximately normally distributed, construct a 90% confidence interval for the ratio of variance in spending for males and females, σ_1^2/σ_2^2.

13. The following information was obtained from two independent samples selected from two normally distributed populations with unknown but equal variances.

    - Sample I: 14, 15, 12, 13, 6, 14, 11, 12, 17, 19, 23.
    - Sample II: 16, 18, 12, 20, 15, 19, 15, 22, 20, 18, 23, 12, 20.

    Test whether the difference of the population means is equal to zero or not. Construct a 95% confidence interval for the difference between the population means and interpret.