STAT 513 fa 2020 Lec 04

Comparing two Normal populations

Karl B. Gregory

Comparing two Normal populations

- We now consider two random samples of sizes n_1 and n_2 drawn from two Normal populations: Let

$$\begin{aligned} &X_{11}, \ldots, X_{1n_1} \text{ be a random sample from the Normal}(\mu_1, \sigma_1^2) \text{ distribution} \\ &X_{21}, \ldots, X_{2n_2} \text{ be a random sample from the Normal}(\mu_2, \sigma_2^2) \text{ distribution,} \end{aligned}$$

  where μ_1, μ_2, σ_1^2, and σ_2^2 are unknown, and let $\bar{X}_1$ and S_1^2 denote the mean and variance of the first sample and $\bar{X}_2$ and S_2^2 denote the mean and variance of the second sample.

- We are interested in

  - comparing μ_1 and μ_2 by testing hypotheses concerning the difference $\mu_1 - \mu_2$ and
  - comparing σ_1^2 and σ_2^2 by testing hypotheses concerning the ratio σ_1^2/σ_2^2

  based on the random samples $X_{11}, \ldots, X_{1n_1}$ and $X_{21}, \ldots, X_{2n_2}$.

Comparing means

- It is simpler to test hypotheses about $\mu_1 - \mu_2$ if $\sigma_1^2 = \sigma_2^2$ than if $\sigma_1^2 \neq \sigma_2^2$, so we begin with the equal-variances case.

- If it is assumed that $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we may pool the data from both samples to estimate the common variance σ^2. The pooled estimator of σ^2 is

$$S^2_{\text{pooled}} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

- We will denote by δ the true value of the difference $\mu_1 - \mu_2$.

- For some value δ_0, which we will call the null value of δ, we consider for some C_1 and C_2 the following tests of hypotheses:

1. *Right-tailed test:* Test H_0: $\mu_1 - \mu_2 \leq \delta_0$ versus H_1: $\mu_1 - \mu_2 > \delta_0$ with the test

$$\text{Reject } H_0 \text{ iff } \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_{\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > C_1.$$

2. *Left-tailed test:* Test H_0: $\mu_1 - \mu_2 \geq \delta_0$ versus H_1: $\mu_1 - \mu_2 < \delta_0$ with the test

$$\text{Reject } H_0 \text{ iff } \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_{\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -C_1.$$

3. *Two-sided test:* Test H_0: $\mu_1 - \mu_2 = \delta_0$ versus H_1: $\mu_1 - \mu_2 \neq \delta_0$ with the test

$$\text{Reject } H_0 \text{ iff } \left|\frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_{\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right| > C_2.$$

- The null value δ_0 is very often chosen to be zero, so that differences in means of any magnitude are of interest. But one may wish to test whether μ_1 is greater than μ_2 by some amount δ_0.

- These are sometimes called *equal-variance two-sample t-tests*, which comes from the fact that if

$$X_{11}, \ldots, X_{1n_1} \text{ is a random sample from the Normal}(\mu_1, \sigma_1^2) \text{ distribution}$$
$$\text{and } X_{21}, \ldots, X_{2n_2} \text{ is a random sample from the Normal}(\mu_2, \sigma_2^2) \text{ distribution}$$

  and if $\sigma_1^2 = \sigma_2^2$, then

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_{\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}.$$

- **Exercise:** *Suppose researchers wish to test, for two populations of interest which are assumed to be Normal and to have the same variance, whether the mean μ_1 of the first population exceeds the mean μ_2 of the second population. They decide to collect a sample of size $n_1 = 20$ from the first population and a sample of size $n_2 = 25$ from the second population and to use the first test with $\delta_0 = 0$ and $C_1 = 2$.*

  (i) *Suppose the means of the two populations are the same. With what probability will the researchers commit a Type I error?*

  (ii) *What would be the effect on this probability of using a smaller value of C_1?*

  (iii) *Suppose the researchers are interested in testing whether the means of the two populations are different, irrespective of which one is greater. Give a test which has size equal to 0.01.*

Answers:

(i) The probability of rejecting H_0 if $\mu_1 = \mu_2$ is

$$P_{\mu_1-\mu_2=0}\left(\frac{\bar{X}_1 - \bar{X}_2 - 0}{S_{\text{pooled}}\sqrt{\frac{1}{20}+\frac{1}{25}}} > 2\right) = P(T > 2), \quad T \sim t_{43}$$
$$= 1 - F_{t_{43}}(2)$$
$$= 1 \text{ - } \texttt{pt(2,43)}$$
$$= 0.02592102.$$

(ii) If we use a smaller value of C_1, the test will reject H_0 based on weaker evidence, so the probability of a Type I error will increase.

(iii) To test H_0: $\mu_1 - \mu_2 = 0$ versus H_1: $\mu_1 - \mu_2 \neq 0$, we will use the two-sided test. To make the test have size 0.01, we find a value of C_2 which satisfies

$$0.01 = P_{\mu_1-\mu_2=0}\left(\left|\frac{\bar{X}_1 - \bar{X}_2 - 0}{S_{\text{pooled}}\sqrt{\frac{1}{20}+\frac{1}{25}}}\right| > C_2\right) = P(|T| > C_2), \quad T \sim t_{43}.$$

We see that the test will have size 0.01 if we choose $C_2 = t_{43,0.005} = \texttt{qt(.995,43)} = 2.695102$.

- If we wish to compute the power of the two-sample t-test for any value of $\mu_1 - \mu_2 = \delta$, we need to make use of the non-central t-distribution.

- We will base power calculations for the equal-variances two-sample t-test on the following result:

Result: Let

$$\begin{aligned} &X_{11}, \dots, X_{1n_1} \text{ be a random sample from the Normal}(\mu_1, \sigma_1^2) \text{ distribution} \\ \text{and } &X_{21}, \dots, X_{2n_2} \text{ be a random sample from the Normal}(\mu_2, \sigma_2^2) \text{ distribution,} \end{aligned}$$

and suppose $\sigma_1^2 = \sigma_2^2$. If $\mu_1 - \mu_2 = \delta$, then

$$\frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_{\text{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{\phi, n_1+n_2-2}, \qquad \text{where } \phi = \frac{\delta - \delta_0}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}.$$

Derivation: We can see this by writing

$$\frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_{\text{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \left(\frac{\bar{X}_1 - \bar{X}_2 - \delta}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} + \frac{\delta - \delta_0}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)\left[\frac{(n_1+n_2-2)S^2_{\text{pooled}}}{\sigma^2}\Big/(n_1+n_2-2)\right]^{-1/2},$$

which has the form

$$\frac{Z + \phi}{\sqrt{W/\nu}},$$

where Z and W are independent random variables such that $Z \sim \text{Normal}(0,1)$ and $W \sim \chi^2_\nu$, and ϕ is a constant; this is the anatomy of a $t_{\phi,\nu}$-distributed random variable.

- **Exercise:** *Get expressions for the power functions in terms of $\delta = \mu_1 - \mu_2$ and the sizes of the right-tailed, left-tailed, and two-sided equal-variance two-sample t-tests. Moreover, for any $\alpha \in (0,1)$, give the values C_1 and C_2 such that the tests have size α.*

  Answer:

  1. The power function for the right-tailed equal-variances two-sample t-test is

     $$\begin{aligned}\gamma(\delta) &= P_{\mu_1-\mu_2=\delta}\left(\frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_{\text{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > C_1\right) \\ &= P(T > C_1), \quad T \sim t_{\phi, n_1+n_2-2}, \quad \phi = \frac{\delta - \delta_0}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \\ &= 1 - F_{t_{\phi,n_1+n_2-2}}(C_1).\end{aligned}$$

     The size of the test is given by the power at $\delta = \delta_0$, for which $\phi = 0$. So the size is

     $$\begin{aligned}\gamma(\delta_0) &= P_{\mu_1-\mu_2=\delta_0}\left(\frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_{\text{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > C_1\right) \\ &= P(T > C_1), \quad T \sim t_{n_1+n_2-2} \\ &= 1 - F_{t_{n_1+n_2-2}}(C_1).\end{aligned}$$

     The test will have size α if $C_1 = t_{n_1+n_2-2,\alpha}$.

  2. The power function for the left-tailed equal-variances two-sample t-test is

     $$\begin{aligned}\gamma(\delta) &= P_{\mu_1-\mu_2=\delta}\left(\frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_{\text{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < -C_1\right) \\ &= P(T < -C_1), \quad T \sim t_{\phi, n_1+n_2-2}, \quad \phi = \frac{\delta - \delta_0}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \\ &= F_{t_{\phi,n_1+n_2-2}}(-C_1).\end{aligned}$$

     The size of the test is given by the power at $\delta = \delta_0$, for which $\phi = 0$. So the size is

     $$\begin{aligned}\gamma(\delta_0) &= P_{\mu_1-\mu_2=\delta_0}\left(\frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{S_{\text{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < -C_1\right) \\ &= P(T < -C_1), \quad T \sim t_{n_1+n_2-2} \\ &= F_{t_{n_1+n_2-2}}(-C_1).\end{aligned}$$

     The test will have size α if $C_1 = t_{n_1+n_2-2,\alpha}$.

3. The power function for the two-sided equal-variances two-sample t-test is

$$\begin{aligned}\gamma(\delta) &= P_{\mu_1-\mu_2=\delta}\left(\left|\frac{\bar{X}_1-\bar{X}_2-\delta_0}{S_{\text{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right| > C_2\right)\\ &= 1-P(-C_2<T<C_2), \quad T\sim t_{\phi,n_1+n_2-2}, \quad \phi = \frac{\delta-\delta_0}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\\ &= 1-[F_{t_{\phi,n_1+n_2-2}}(C_2)-F_{t_{\phi,n_1+n_2-2}}(-C_2)].\end{aligned}$$

The size of the test is given by the power at $\delta=\delta_0$, for which $\phi=0$. So the size is

$$\begin{aligned}\gamma(\delta_0) &= P_{\mu_1-\mu_2=\delta_0}\left(\left|\frac{\bar{X}_1-\bar{X}_2-\delta_0}{S_{\text{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right| > C_2\right)\\ &= 1-P(-C_2<T<C_2), \quad T\sim t_{n_1+n_2-2}\\ &= 1-[F_{t_{n_1+n_2-2}}(C_2)-F_{t_{n_1+n_2-2}}(-C_2)].\end{aligned}$$

The test will have size α if $C_2 = t_{n_1+n_2-2,\alpha/2}$.

- **Exercise:** *Let*

$$X_{11},\dots,X_{1n_1} \text{ be a random sample from the Normal}(\mu_1,\sigma_1^2)\text{ distribution}$$
$$\text{and } X_{21},\dots,X_{2n_2} \text{ be a random sample from the Normal}(\mu_2,\sigma_2^2)\text{ distribution}$$

with $\sigma_1^2=\sigma_2^2$, *and suppose there are three researchers:*

  - *Simone to test* H_0: $\mu_1-\mu_2\leq 0$ *vs* H_1: $\mu_1-\mu_2>0$ *with Rej.* H_0 *iff* $\frac{\bar{X}_1-\bar{X}_2}{S_{\text{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > t_{n_1+n_2-2,0.10}$
  - *Helmut to test* H_0: $\mu_1-\mu_2\geq 0$ *vs* H_1: $\mu_1-\mu_2<0$ *with Rej.* H_0 *iff* $\frac{\bar{X}_1-\bar{X}_2}{S_{\text{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < -t_{n_1+n_2-2,0.10}$
  - *Mechthild to test* H_0: $\mu_1-\mu_2=0$ *vs* H_1: $\mu_1-\mu_2\neq 0$ *with Rej.* H_0 *iff* $\left|\frac{\bar{X}_1-\bar{X}_2}{S_{\text{pooled}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right| > t_{n_1+n_2-2,0.05}$

(i) *What is the size of each test?*

(ii) *Suppose* $\mu_1=1$ *and* $\mu_2=2$. *Which researcher is most likely to reject* H_0*?*

(iii) *Suppose* $\mu_1=1$ *and* $\mu_2=2$. *Which researcher is least likely to reject* H_0*?*

(iv) *Suppose* $\mu_1=1$ *and* $\mu_2=2$. *Which researcher may commit a Type I error?*

(v) *Suppose* $\mu_1=2$ *and* $\mu_2=1$. *Which researcher may commit a Type I error?*

(vi) *Suppose* $\mu_1=1$ *and* $\mu_2=2$. *Which researchers may commit a Type II error?*

(vii) *Suppose* $\mu_1=2$ *and* $\mu_2=1$. *Which researchers may commit a Type II error?*

(viii) *Based on the sample sizes* $n_1=10$ *and* $n_2=15$, *compute the power of each researcher's test when* $\mu_1=2$ *and* $\mu_2=1$ *and* $\sigma_1^2=\sigma_2^2=\sigma^2=4$.

(ix) *Plot the power curves of the three researchers' tests together, where power is a function of* $\delta = \mu_1 - \mu_2$.

(x) *Suppose that with sample sizes* $n_1 = 10$ *and* $n_2 = 15$, $\bar{X}_1 = 0.97$, $\bar{X}_2 = 1.26$, $S_1 = 0.24$, *and* $S_2 = 0.19$. *Compute the* p*-value of each researcher's test.*

Answers:

(viii) If $\mu_1 = 2$ and $\mu_2 = 1$ then $\delta = \mu_1 - \mu_2 = 1$. In the following, let $T \sim t_{\phi,23}$ (note that $10 + 15 - 2 = 23$), where $\phi = (1-0)/(2\sqrt{1/10 + 1/15})$.

- The test of Simone has power

$$\begin{aligned} 1 - P(T > t_{23,0.10}) &= 1 - F_{t_{\phi,23}}(t_{23,0.10}) \\ &= \texttt{1 - pt(qt(.9,23),23,1/(2*sqrt(1/10+1/15)))} \\ &= 0.4685883. \end{aligned}$$

- The test of Helmut has power

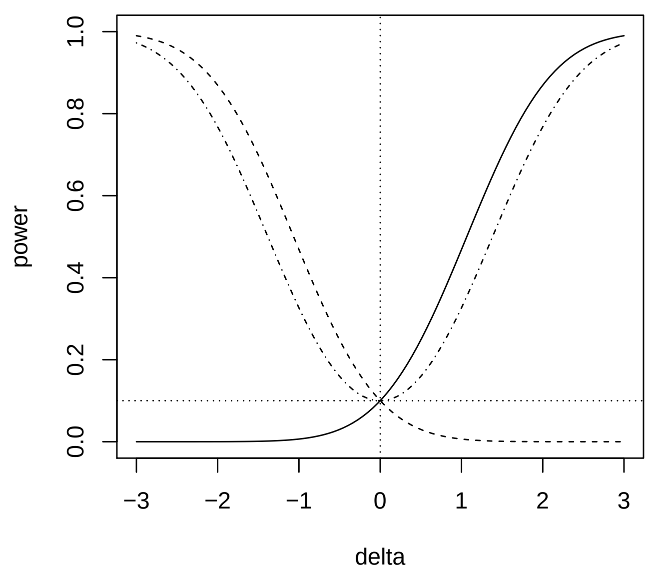
$$\begin{aligned} P(T < -t_{23,0.10}) &= F_{t_{\phi,23}}(-t_{23,0.10}) \\ &= \texttt{pt(-qt(.9,23),23,1/(2*sqrt(1/10+1/15)))} \\ &= 0.006481496. \end{aligned}$$

- The test of Mechthild has power

$$\begin{aligned} 1 - P(|T| > t_{23,0.05}) &= 1 - [F_{t_{\phi,23}}(t_{23,0.05}) - F_{t_{\phi,23}}(-t_{23,0.05})] \\ &= \texttt{1 - (pt(qt(.95,23),23,1/(2*sqrt(1/10+1/15)))} \\ &\qquad\qquad \texttt{-pt(-qt(.95,23),23,1/(2*sqrt(1/10+1/15))))} \\ &= 0.3264441. \end{aligned}$$

(ix) The following R code makes the plot:

```
delta.seq <- seq(-3,3,length=100)
delta.0 <- 0
n1 <- 10
n2 <- 15
alpha <- 0.10
sigma <- 2

phi <- (delta.seq-delta.0)/(sigma*sqrt(1/n1+1/n2))

power.gt <- 1 - pt(qt(1-alpha,n1+n2-2),n1+n2-2,phi)
power.lt <- pt(-qt(1-alpha,n1+n2-2),n1+n2-2,phi)
power.neq <- 1-(pt(qt(1-alpha/2,n1+n2-2),n1+n2-2,phi)
                - pt(-qt(1-alpha/2,n1+n2-2),n1+n2-2,phi))

plot(delta.seq, power.gt,type="l",ylim=c(0,1),xlab="delta",ylab="power")
lines(delta.seq, power.lt,lty=2)
lines(delta.seq, power.neq,lty=4)

abline(v=delta.0,lty=3) # vert line at null value
abline(h=0.10,lty=3)         # horiz line at size
```



(x) The estimate of the common variance σ^2 is

$$S^2_{\text{pooled}} = \frac{(10-1)(0.24)^2 + (15-1)(0.19)^2}{10+15-2} = \texttt{(9*0.24**2+14*0.19**2)/23} = 0.04451304,$$

so that we get

$$\frac{\bar{X}_1 - \bar{X}_2}{S_{\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0.97 - 1.26}{\sqrt{0.04451304}\sqrt{1/10 + 1/15}}$$

$= \;$ `(0.97-1.26)/sqrt(0.04451304*(1/10+1/15))`

$= -3.366898.$

So the p-value for

- Simone's test is $1 - F_{t_{23}}(-3.366898) =$ `1-pt(-3.366898,23)` $= 0.9986682$.
- Helmut's test is $F_{t_{23}}(-3.366898) =$ `pt(-3.366898,23)` $= 0.001331817$.
- Mechthild's test is $2(1 - F_{t_{23}}(|-3.366898|)) =$ `2*(1-pt(3.366898,23))` $= 0.002663634$.

- In the unequal-variances case, that is when $\sigma_2^2 \neq \sigma_2^2$, we cannot pool the data together from the two samples to estimate a common variance; we must estimate the two variances separately. As a result, we use a different test statistic, the null distribution (the distribution when $\mu_1 - \mu_2 = \delta_0$) of which is approximated by a t-distribution with a certain—complicated—degrees of freedom value. This has been discussed in STAT 512, and we will not discuss this case in detail here. The following formulas for two-sample t-tests specify how to perform both the equal- and unequal-variances two-sample t-tests.

- **Formulas for two-sample t-tests:** Let

$$\begin{aligned} &X_{11}, \ldots, X_{1n_1} \text{ be a random sample from the Normal}(\mu_1, \sigma_1^2) \text{ distribution} \\ \text{and } &X_{21}, \ldots, X_{2n_2} \text{ be a random sample from the Normal}(\mu_2, \sigma_2^2) \text{ distribution,} \end{aligned}$$

and for some δ_0, define

$$T = \begin{cases} (\bar{X}_1 - \bar{X}_2 - \delta_0)/(S_{\text{pooled}}\sqrt{1/n_1 + 1/n_2}), & \text{if believed } \sigma_1^2 = \sigma_2^2 = \sigma^2 \\ (\bar{X}_1 - \bar{X}_2 - \delta_0)/\sqrt{S_1^2/n_1 + S_2^2/n_2}, & \text{if believed } \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

and

$$\nu = \begin{cases} n_1 + n_2 - 2, & \text{if believed } \sigma_1^2 = \sigma_2^2 = \sigma^2 \\ \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2 \left[\left(\frac{S_1}{n_1}\right)^2/(n_1 - 1) + \left(\frac{S_2^2}{n_2}\right)^2/(n_2 - 1)\right]^{-1}, & \text{if believed } \sigma_1^2 \neq \sigma_2^2. \end{cases}$$

Then we have the following, where $F_{t_{\phi,\nu}}$ is the cdf of the $t_{\phi,\nu}$-distribution:

H_1	Reject H_0 at α iff	Power function $\gamma(\mu)$	p-value
$\mu_1 - \mu_2 > \delta_0$	$T > t_{\nu,\alpha}$	$1 - F_{t_{\phi,\nu}}(t_{\nu,\alpha})$	$1 - F_{t_\nu}(T)$
$\mu_1 - \mu_2 < \delta_0$	$T < -t_{\nu,\alpha}$	$F_{t_{\phi,\nu}}(-t_{\nu,\alpha})$	$F_{t_\nu}(T)$
$\mu_1 - \mu_2 \neq \delta_0$	$\|T\| > t_{\nu,\alpha/2}$	$1 - [F_{t_{\phi,\nu}}(t_{\nu,\alpha/2}) - F_{t_{\phi,\nu}}(-t_{\nu,\alpha/2})]$	$2[1 - F_{t_\nu}(\|T\|)]$,

where

$$\phi = \begin{cases} (\delta - \delta_0)/(\sigma\sqrt{1/n_1 + 1/n_2}), & \text{if believed } \sigma_1^2 = \sigma_2^2 = \sigma^2 \\ (\delta - \delta_0)/(\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}), & \text{if believed } \sigma_1^2 \neq \sigma_2^2. \end{cases}$$

Comparing variances

- It is easier to get a handle on the distribution of the ratio S_1^2/S_2^2 than on that of the difference $S_1^2 - S_2^2$, so we make comparisons of σ_1^2 and σ_2^2 in terms of the ratio σ_1^2/σ_2^2.
- We will denote by ϑ the true value of the ratio σ_1^2/σ_2^2.
- For some value ϑ_0, which we will call the null value of ϑ, we consider for some C_{1r}, C_{1l}, C_{2l}, and C_{2r} the following tests of hypotheses:

  1. *Right-tailed test:* Test H_0: $\sigma_1^2/\sigma_2^2 \leq \vartheta_0$ versus H_1: $\sigma_1^2/\sigma_2^2 > \vartheta_0$ with the test

  $$\text{Reject } H_0 \text{ iff } \frac{S_1^2}{S_2^2}/\vartheta_0 > C_{1r}.$$

  2. *Left-tailed test:* Test H_0: $\sigma_1^2/\sigma_2^2 \geq \vartheta_0$ versus H_1: $\sigma_1^2/\sigma_2^2 < \vartheta_0$ with the test

  $$\text{Reject } H_0 \text{ iff } \frac{S_1^2}{S_2^2}/\vartheta_0 < C_{1l}.$$

  3. *Two-sided test:* Test H_0: $\sigma_1^2/\sigma_2^2 = \vartheta_0$ versus H_1: $\sigma_1^2/\sigma_2^2 \neq \vartheta_0$ with the test

  $$\text{Reject } H_0 \text{ iff } \frac{S_1^2}{S_2^2}/\vartheta_0 < C_{2l} \text{ or } \frac{S_1^2}{S_2^2}/\vartheta_0 > C_{2r}.$$

- The null ratio ϑ_0 is very often chosen as $\vartheta_0 = 1$, in which case it is of interest whether there is a difference of any magnitude between σ_1^2 and σ_2^2. However, if one wishes to test, for example, whether $\sigma 1^2$ is more than twice as large as σ_2^2, one may choose $\vartheta_0 = 2$.
- We will call these tests *variance ratio F-tests*, which comes from the fact that if

  $$X_{11}, \ldots, X_{1n_1} \text{ is a random sample from the Normal}(\mu_1, \sigma_1^2) \text{ distribution}$$
  $$\text{and } X_{21}, \ldots, X_{2n_2} \text{ is a random sample from the Normal}(\mu_2, \sigma_2^2) \text{ distribution}$$

  and if $\sigma_1^2/\sigma_2^2 = \vartheta$, then

  $$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2}/\vartheta \sim F_{n_1-1,n_2-1},$$

  where F_{n_1-1,n_2-1} denotes the F-distribution with numerator degrees of freedom $n_1 - 1$ and denominator degrees of freedom $n_2 - 1$.

- For $\xi \in (0,1)$, let $F_{n_1-1,n_2-1,\xi}$ be the value which satisfies $\xi = P(R > F_{n_1-1,n_2-1,\xi})$, when $R \sim F_{n_1-1,n_2-1}$.
- Let $F_{F_{n_1-1,n_2-1}}$ denote the cdf of the F_{n_1-1,n_2-1}-distribution.
- **Exercise:** *Suppose researchers wish to test, for two populations of interest which are assumed to be Normal, whether the variance σ_1^2 of the first population is equal to the variance σ_2^2 of the second population. They decide to collect a sample of size $n_1 = 20$ from the first population and a sample of size $n_2 = 25$ from the second population and to use the two-sided test with $\vartheta_0 = 1$ and $C_{2l} = 1/2$ and $C_{2r} = 2$; that is, they will reject H_0: $\sigma_1^2 = \sigma_2^2$ iff $S_1^2/S_2^2 < 1/2$ or $S_1^2/S_2^2 > 2$.*

  (i) What is the size of the test?

  (ii) If $\sigma_1^2 = 1.5$ and $\sigma_2^2 = 1$, which type of inferential error is possible?

  (iii) If $\sigma_1^2 = 1.5$ and $\sigma_2^2 = 1$, what is the probability the the test will lead to an incorrect decision?

  Answers:

  (i) The size of the test is

$$\begin{aligned} P_{\sigma_1^2/\sigma_2^2=1}(S_1^2/S_2^2 < 1/2) + P_{\sigma_1^2/\sigma_2^2=1}(S_1^2/S_2^2 > 2) &= P(R < 1/2) + P(R > 2), \quad R \sim F_{19,24} \\ &= \texttt{pf(1/2,19,24)+1-pf(2,19,24)} \\ &= 0.1183526. \end{aligned}$$

  (ii) If $\sigma_1^2 = 1.5$ and $\sigma_2^2 = 1$, the correct decision is to reject H_0; if we fail to reject it, it is a Type II error.

  (iii) If $\sigma_1^2 = 1.5$ and $\sigma_2^2 = 1$, the probability of a Type II error is

$$\begin{aligned} P_{\sigma_1^2/\sigma_2^2=3/2}(1/2 < S_1^2/S_2^2 < 2) &= P_{\sigma_1^2/\sigma_2^2=3/2}((1/2)/(3/2) < (S_1^2/S_2^2)/(3/2) < 2/(3/2)) \\ &= P(1/3 < R < 4/3), \quad R \sim F_{19,24} \\ &= \texttt{pf(4/3,19,24)-pf(1/3,19,24)} \\ &= 0.7412179. \end{aligned}$$

- **Exercise:** *Get expressions for the power functions in terms of the true ratio of the variances $\vartheta = \sigma_1^2/\sigma_2^2$ of the right-tailed, left-tailed, and two-sided variance ratio F-tests and calibrate the rejection regions, that is, find C_{1l}, C_{1r}, C_{2l}, and C_{2r}, such that for any $\alpha \in (0,1)$ each test has size α.*

Answers:

1. The right-tailed test has power given by

$$\begin{aligned}\gamma(\vartheta) &= P_{\sigma_1^2/\sigma_2^2=\vartheta}((S_1^2/S_2^2)/\vartheta_0 > C_{1r})\\ &= P_{\sigma_1^2/\sigma_2^2=\vartheta}((S_1^2/S_2^2)/\vartheta > C_{1r}(\vartheta_0/\vartheta))\\ &= P(R > C_{1r}(\vartheta_0/\vartheta)), \quad R \sim F_{n_1-1,n_2-1}\\ &= 1 - F_{F_{n_1-1,n_2-1}}(C_{1r}(\vartheta_0/\vartheta)).\end{aligned}$$

   The size is given by

$$\gamma(\vartheta_0) = 1 - F_{F_{n_1-1,n_2-1}}(C_{1r}).$$

   The test will have size equal to α if $C_{1r} = F_{n_1-1,n_2-1,\alpha}$.

2. The left-tailed test has power given by

$$\begin{aligned}\gamma(\vartheta) &= P_{\sigma_1^2/\sigma_2^2=\vartheta}((S_1^2/S_2^2)/\vartheta_0 < C_{1l})\\ &= P_{\sigma_1^2/\sigma_2^2=\vartheta}((S_1^2/S_2^2)/\vartheta < C_{1l}(\vartheta_0/\vartheta))\\ &= P(R < C_{1l}(\vartheta_0/\vartheta)), \quad R \sim F_{n_1-1,n_2-1}\\ &= F_{F_{n_1-1,n_2-1}}(C_{1l}(\vartheta_0/\vartheta)).\end{aligned}$$

   The size is given by

$$\gamma(\vartheta_0) = F_{F_{n_1-1,n_2-1}}(C_{1l}).$$

   The test will have size equal to α if $C_{1l} = F_{n_1-1,n_2-1,1-\alpha}$.

3. The two-sided test has power given by

$$\begin{aligned}\gamma(\vartheta) &= P_{\sigma_1^2/\sigma_2^2=\vartheta}((S_1^2/S_2^2)/\vartheta_0 < C_{2l}) + P_{\sigma_1^2/\sigma_2^2=\vartheta}((S_1^2/S_2^2)/\vartheta_0 > C_{2r})\\ &= P_{\sigma_1^2/\sigma_2^2=\vartheta}((S_1^2/S_2^2)/\vartheta < C_{2l}(\vartheta_0/\vartheta)) + P_{\sigma_1^2/\sigma_2^2=\vartheta}((S_1^2/S_2^2)/\vartheta > C_{2r}(\vartheta_0/\vartheta))\\ &= P(R < C_{2l}(\vartheta_0/\vartheta)) + P(R > C_{2r}(\vartheta_0/\vartheta)), \quad R \sim F_{n_1-1,n_2-1}\\ &= F_{F_{n_1-1,n_2-1}}(C_{1l}(\vartheta_0/\vartheta)) + 1 - F_{F_{n_1-1,n_2-1}}(C_{1r}(\vartheta_0/\vartheta)).\end{aligned}$$

   The size is given by

$$\gamma(\vartheta_0) = F_{F_{n_1-1,n_2-1}}(C_{1l}) + 1 - F_{F_{n_1-1,n_2-1}}(C_{1r}).$$

   The test will have size equal to α if $C_{2l} = F_{n_1-1,n_2-1,1-\alpha/2}$ and $C_{2r} = F_{n_1-1,n_2-1,\alpha/2}$.

- **Exercise:** *Let*

$$\begin{aligned}&X_{11},\ldots,X_{1n_1} \textit{ be a random sample from the Normal}(\mu_1,\sigma_1^2) \textit{ distribution}\\ \textit{and } &X_{21},\ldots,X_{2n_2} \textit{ be a random sample from the Normal}(\mu_2,\sigma_2^2) \textit{ distribution,}\end{aligned}$$

and suppose there are three researchers:

  - *Kwame to test* H_0*:* $\sigma_1^2/\sigma_2^2 \leq 1$ *vs* H_1*:* $\sigma_1^2/\sigma_2^2 > 1$ *with Rej.* H_0 *iff* $S_1^2/S_2^2 > F_{n_1-1,n_2-1,0.10}$

- *Ama to test* H_0*:* $\sigma_1^2/\sigma_2^2 \geq 1$ *vs* H_1*:* $\sigma_1^2/\sigma_2^2 < 1$ *with Rej.* H_0 *iff* $S_1^2/S_2^2 < F_{n_1-1,n_2-1,0.90}$
- *Kobe to test* H_0*:* $\sigma_1^2/\sigma_2^2 = 1$ *vs* H_1*:* $\sigma_1^2/\sigma_2^2 \neq 1$ *with Rej.* H_0 *iff* $S_1^2/S_2^2 < F_{n_1-1,n_2-1,0.95}$ *or* $S_1^2/S_2^2 > F_{n_1-1,n_2-1,0.05}$

(i) *What is the size of each test?*

(ii) *Suppose* $\sigma_1^2 = 1$ *and* $\sigma_2^2 = 2$*. Which researcher is most likely to reject* H_0*?*

(iii) *Suppose* $\sigma_1^2 = 1$ *and* $\sigma_2^2 = 2$*. Which researcher is least likely to reject* H_0*?*

(iv) *Suppose* $\sigma_1^2 = 1$ *and* $\sigma_2^2 = 2$*. Which researcher may commit a Type I error?*

(v) *Suppose* $\sigma_1^2 = 2$ *and* $\sigma_2^2 = 1$*. Which researcher may commit a Type I error?*

(vi) *Suppose* $\sigma_1^2 = 1$ *and* $\sigma_2^2 = 2$*. Which researchers may commit a Type II error?*

(vii) *Suppose* $\sigma_1^2 = 2$ *and* $\sigma_2^2 = 1$*. Which researchers may commit a Type II error?*

(viii) *Based on the sample sizes* $n_1 = 18$ *and* $n_2 = 17$*, compute the power of each researcher's test when* $\sigma_1^2 = 1$ *and* $\sigma_2^2 = 2$*.*

(ix) *Plot the power curves of the three researchers' tests together, where power is a function of* $\vartheta = \sigma_1^2/\sigma_2^2$*.*

(x) *Suppose that with samples of sizes* $n_1 = 18$ *and* $n_2 = 17$*,* $S_1^2 = 1.15$ *and* $S_2^2 = 1.87$ *is observed. Compute the* p*-value of each researcher's test.*

Answers:

(viii) We have $\vartheta_0 = 1$, and if $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$ then $\vartheta = \sigma_1^2/\sigma_2^2 = 2$. In the following, let $R \sim F_{17,16}$.

- The test of Kwame has power

$$\begin{aligned} P(R > F_{17,16,0.10}(1/2)) &= 1 - F_{F_{17,16}}(F_{17,16,0.10}/2) \\ &= 1 \texttt{ - pf(qf(.9,17,16)/2,17,16)} \\ &= 0.5355808. \end{aligned}$$

- The test of Ama has power

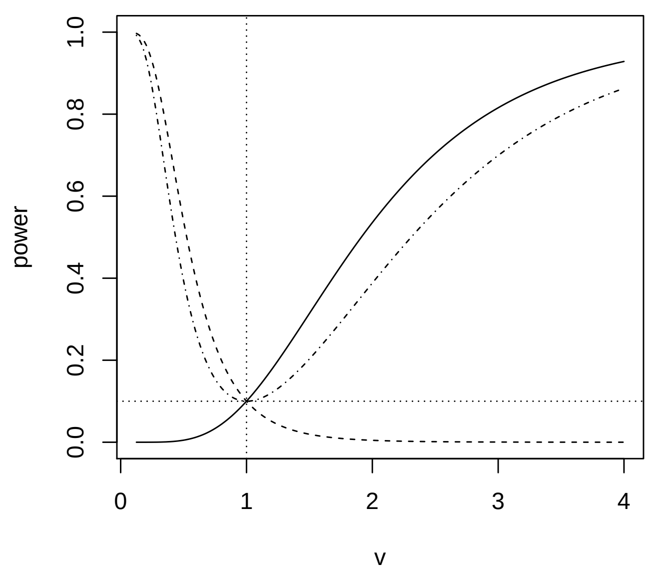
$$\begin{aligned} P(R < F_{17,16,0.90}(1/2)) &= F_{F_{17,16}}(F_{17,16,0.90}/2) \\ &= \texttt{pf(qf(.10,17,16)/2,17,16)} \\ &= 0.004645553. \end{aligned}$$

- The test of Kobe has power

$$\begin{aligned} &P(R < F_{17,16,0.95}(1/2)) + P(R > F_{17,16,0.05}/2) \\ &\quad = F_{F_{17,16}}(F_{17,16,0.95}/2) + 1 - F_{F_{17,16}}(F_{17,16,0.05}/2) \\ &\quad = \texttt{pf(qf(.05,17,16)/2,17,16)} \\ &\qquad\qquad \texttt{+1-pf(qf(.95,17,16)/2,17,16)} \\ &\quad = 0.3880993. \end{aligned}$$

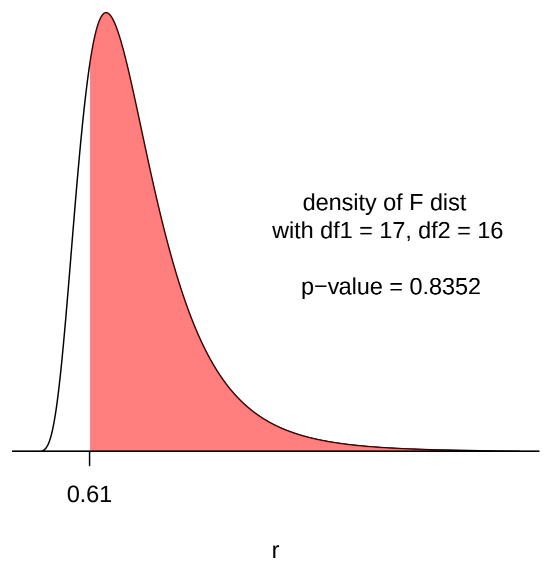
(ix) The following R code makes the plot:

```
v.seq <- seq(1/8,4,length=200)
v.0 <- 1
n1 <- 18
n2 <- 17
alpha <- 0.10

power.gt <- 1 - pf(qf(1-alpha,n1-1,n2-1)*v.0/v.seq,n1-1,n2-1)
power.lt <- pf(qf(alpha,n1-1,n2-1)*v.0/v.seq,n1-1,n2-1)
power.neq <- pf(qf(alpha/2,n1-1,n2-1)*v.0/v.seq,n1-1,n2-1)
                + 1 - pf(qf(1-alpha/2,n1-1,n2-1)*v.0/v.seq,n1-1,n2-1)

plot(v.seq, power.gt,type="l",ylim=c(0,1),xlab="v",ylab="power")
lines(v.seq, power.lt,lty=2)
lines(v.seq, power.neq,lty=4)

abline(v=v.0,lty=3)  # vert line at null value
abline(h=0.10,lty=3) # horiz line at size
```

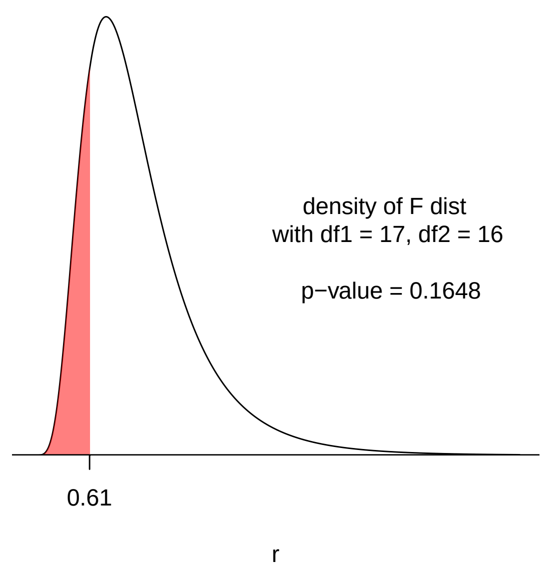


(x) We have $S_1^2/S_2^2 = 1.15/1.87 = 0.6149733$. So the p-value

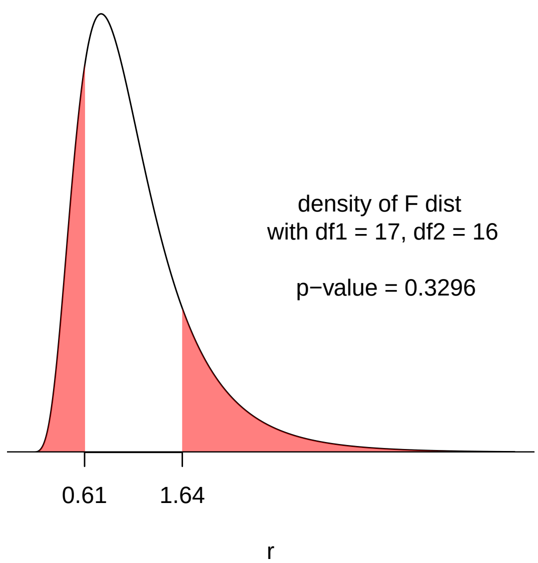
- for Kwame's test is $1 - F_{F_{17,16}}(0.6149733) =$ `1-pf(0.6149733,17,16)` $= 0.8351805$.



- for Ama's test is $F_{F_{17,16}}(0.6149733) =$ `pf(0.6149733,17,16)` $= 0.1648195$.



– for Kobe's test is $2[F_{F_{17,16}}(0.6149733)] =$ `2*pf(0.6149733,17,16)` $= 0.329639$.



- **Formulas for variance ratio F-tests:** Let

$$X_{11}, \ldots, X_{1n_1} \text{ be a random sample from the Normal}(\mu_1, \sigma_1^2) \text{ distribution}$$
$$\text{and } X_{21}, \ldots, X_{2n_2} \text{ be a random sample from the Normal}(\mu_2, \sigma_2^2) \text{ distribution,}$$

and for some ϑ_0, define $R = (S_1^2/S_2^2)/\vartheta_0$. Then we have the following:

H_1	Reject H_0 at α iff	Power function $\gamma(\vartheta)$	p-value
$\sigma_1^2/\sigma_2^2 > \vartheta_0$	$R > F_{n_1-1,n_2-1,\alpha}$	$1 - F_{F_{n_1-1,n_2-1}}(F_{n_1-1,n_2-1,\alpha}(\vartheta_0/\vartheta))$	$1 - F_{F_{n_1-1,n_2-1}}(R)$
$\sigma_1^2/\sigma_2^2 < \vartheta_0$	$R < F_{n_1-1,n_2-1,1-\alpha}$	$F_{F_{n_1-1,n_2-1}}(F_{n_1-1,n_2-1,1-\alpha}(\vartheta_0/\vartheta))$	$F_{F_{n_1-1,n_2-1}}(R)$
$\sigma_1^2/\sigma_2^2 \neq \vartheta_0$	$R < F_{n_1-1,n_2-1,1-\alpha/2}$ or $R > F_{n_1-1,n_2-1,\alpha/2}$	$F_{F_{n_1-1,n_2-1}}(F_{n_1-1,n_2-1,1-\alpha/2}(\vartheta_0/\vartheta))$ $+ 1 - F_{F_{n_1-1,n_2-1}}(F_{n_1-1,n_2-1,\alpha/2}(\vartheta_0/\vartheta))$	$2 \cdot \min\{F_{F_{n_1-1,n_2-1}}(R),$ $1 - F_{F_{n_1-1,n_2-1}}(R)\}$