# Solutions to HW Problems in Chapter 7

7.1 Begin with

$$B(d) = R(d;\zeta_1)P_1 + R(d;\zeta_2)P_2$$
$$= \left(l(a_1;\zeta_1)P[a_1 \mid \zeta_1] + l(a_2;\zeta_1)P[a_2 \mid \zeta_1]\right)P_1$$
$$+\left(l(a_1;\zeta_2)P[a_1 \mid \zeta_2] + l(a_2;\zeta_2)P[a_2 \mid \zeta_2]\right)P_2.$$

Next observe that

$$P[a_1 \mid \zeta_2] = \int_c^\infty f(x;\zeta_2)dx;\ P[a_2 \mid \zeta_2] = \int_{-\infty}^c f(x;\zeta_2)dx = 1 - \int_c^\infty f(x;\zeta_2)dx;$$
$$P[a_1 \mid \zeta_1] = \int_c^\infty f(x;\zeta_1)dx;\ P[a_2 \mid \zeta_1] = \int_{-\infty}^c f(x;\zeta_1)dx = 1 - \int_c^\infty f(x;\zeta_1)dx;$$

Inserting these results in the expression for $B(d)$ yields:

$$B(d)\left[l(a_1;\zeta_1)\int_c^\infty f(x;\zeta_1)dx + l(a_2;\zeta_1)\left(1 - \int_c^\infty f(x;\zeta_1)dx\right)\right]\times P_1$$
$$+\left[l(a_1;\zeta_2)\int_c^\infty f(x;\zeta_2)dx + l(a_2;\zeta_2)\left(1 - \int_c^\infty f(x;\zeta_2)dx\right)\right]\times P_2,$$

whence Eq.(7.1-6) follows.

---

7.2. When $P_1 = 0.9$ and $P_2 = 0.1$ we can expect that cancer is almost always present. We expect the point $c$ to move far t the left so that almost any realization of $X$ will lead the surgeon to operate. Specializing Eq. (7.1-6) for this case, keeping the loss functions the same as in Table 7.1-1, leads to $B(d) = 31.5 + \int_c^\infty \left(0.5f(x;\zeta_2) - 30f(x;\zeta_1)\right)dx$ . The Bays solution is when:

$$0.5f(x;\zeta_2) - 30f(x;\zeta_1) < 0 \text{ or}$$
$$f(c;\zeta_1)/f(c;\zeta_2) = 0.0167 \quad .$$

Thus when the event $\{X > c\}$ occurs, the surgeon should operate.

---

7.3 The test is $(n)^{-1}\sum_{i=1}^{n} X_i \triangleq \hat{\mu} > c$ for rejecting the hypothesis. Here *c* is obtained from

$0.05 = \frac{1}{\sqrt{2\pi\sigma^2 / n}} \int_c^{\infty} \exp\left( \frac{-1}{2} \left( \frac{x-\mu_1}{\sigma/\sqrt{n}} \right)^2 \right) dx$. When $\sigma = 1$, and converting to the standard, Normal, we get

$0.95 = F_{SN}\left( (\sqrt{n}(c-\mu_1) \right) = F_{SN}(1.645)$. Hence $c = \mu_1 + 1.645/\sqrt{n} \to \mu_1$ as $n \to \infty$.

---

7.4 The power P of a test is $\mathrm{P} = P[reject\ H_1 \mid H_2\ true] = 1 - P[accept\ H_1 \mid H_2\ true]$. Note

$$P[H_2\ true] = P[reject\ H_1, H_2\ true] + P[accept\ H_1, H_2\ true]$$
$$= P[reject\ H_1 \mid H_2\ true]P[H_2\ true] + P[accept\ H_1 \mid H_2\ true]P[H_2\ true]$$

or, equivalently,

$$1 = P[reject\ H_1 \mid H_2\ true] + P[accept\ H_1 \mid H_2\ true]$$

from which it follows that

$$\mathrm{P} = 1 - P[accept\ H_1 \mid H_2\ true].$$

---

7.5 In Example 7.2-2 it was assumed that *X* and therefore each of its observation $X_i, i = 1, ..., n$ were $N(\mu, \sigma^2)$. The MGF of a Normal RV is

$M_X = \exp(\mu t) \times \exp(\sigma^2 t^2 / 2)$. From the properties of moment generating functions we have that the moment generating function of a sum of i.i.d. RVs is simply the product of their MGFs. Hence

$Y \triangleq \sum_{i=1}^{n} X_i \leftrightarrow \exp(n\mu t) \times \exp(n\sigma^2 t^2 / 2)$ and *Y* is seen to be $N(n\mu, n\sigma^2)$. Finally, from elementary probability, the transformation $\hat{\mu} \triangleq Y/n$ yields a $N(\mu, \sigma^2 / n)$ RV.

---

7.6 Here we use the Normal approximation to the binomial since *n* is large. Under $H_1$ we find that $\mu = 100 \times 0.5 = 50$ and $\sigma^2 = 100 \times 0.5 \times 0.5 = 25$. Hence we seek the number *c* such that

$$\alpha = \frac{1}{5\sqrt{2\pi}}\int_k^{\infty} exp\left(-0.5\left((x-50)/5\right)^2\right)dx + \frac{1}{5\sqrt{2\pi}}\int_{-\infty}^{-k} exp\left(-0.5\left((x-50)/5\right)^2\right)dx$$ or, equivalently

$$1-\alpha = 2\times erf\,(\frac{k-50}{5})$$

At the 0.05 level we find $k = 50 \pm 10$. Hence if either the number of heads or tails exceeds 60 or is less than 40 , the hypothesis is rejected. At the $\alpha = 0.1$ level, we find that $k = 50 \pm (5\times 1.65)$. Hence if either the number of heads or tails exceeds 58 or is less than 42, the hypothesis is rejected.

---

7.7 The four decision functions are:

1. Buy the battery if it starts the car, else reject the battery;
2. Buy the battery no matter what;
3. Don't buy the battery no matter what;
4. Don't buy the battery if it starts the car, else buy it.

Put into symbols we get

$$d_1(X): d_1(1) = a_1;\ d_1(0) = a_2$$
$$d_2(X): d_1(1) = a_1;\ d_1(0) = a_1$$
$$d_3(X): d_1(1) = a_2;\ d_1(0) = a_2$$
$$d_4(X): d_1(1) = a_2;\ d_1(0) = a_1$$

We denote the by $\begin{cases} \zeta_1 \text{ the outcome that the battery is of the superior type i.e. from A} \\ \zeta_2 \text{ the outcome that the battery is of the inferior type i.e. from B} \end{cases}$

The state of nature $\zeta_1$ corresponds to battery A with start probability $p_1 = 0.8$. For convenience we write $\zeta_1$=0.8. Likewise the state of nature $\zeta_2$ corresponds to battery B with start probability $p_2 = 0.5$. For convenience we write $\zeta_2 = 0.5$.

The loss functions are: $l(a_1,\zeta_1) = 0;\ l(a_1,\zeta_2) = 40;\ l(a_2,\zeta_1) = 10;\ l(a_2,\zeta_2) = 0.$

The risk formula is: $R(d,\zeta) = l(a_1,\zeta)P[a_1 \mid \zeta] + l(a_2,\zeta)P[a_2 \mid \zeta]$.

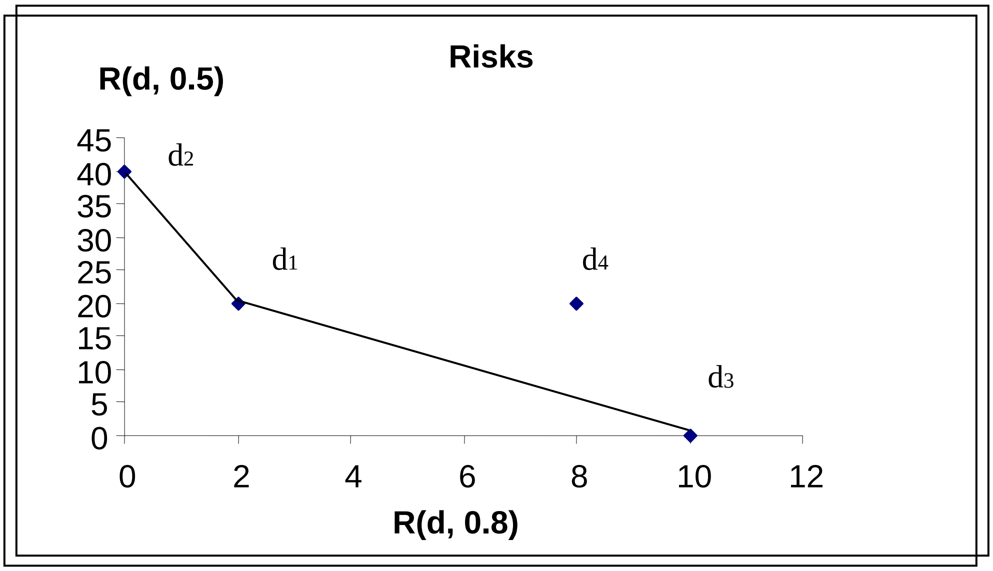
Thus:

$$R(d_1,\zeta_1) = l(a_1,\zeta_1)P[a_1 \mid \zeta_1] + l(a_2,\zeta_1)P[a_2 \mid \zeta_1] = 0 + 10\times 0.2 = 2$$
$$R(d_1,\zeta_2) = l(a_1,\zeta_2)P[a_1 \mid \zeta_2] + l(a_2,\zeta_2)P[a_2 \mid \zeta_2] = 40\times 0.5 = 20$$

$$R(d_2,\zeta_1) = l(a_1,\zeta_1)P[a_1 \mid \zeta_1] + l(a_2,\zeta_1)P[a_2 \mid \zeta_1] = 0$$
$$R(d_2,\zeta_2) = l(a_1,\zeta_2)P[a_1 \mid \zeta_2] + l(a_2,\zeta_2)P[a_2 \mid \zeta_2] = 40$$

$$R(d_3,\zeta_1) = l(a_1,\zeta_1)P[a_1 \mid \zeta_1] + l(a_2,\zeta_1)P[a_2 \mid \zeta_1] = 10$$
$$R(d_3,\zeta_2) = l(a_1,\zeta_2)P[a_1 \mid \zeta_2] + l(a_2,\zeta_2)P[a_2 \mid \zeta_2] = 0$$

$$R(d_4,\zeta_1) = l(a_1,\zeta_1)P[a_1 \mid \zeta_1] + l(a_2,\zeta_1)P[a_2 \mid \zeta_1] = 0 + 10 \times 0.8 = 8$$
$$R(d_4,\zeta_2) = l(a_1,\zeta_2)P[a_1 \mid \zeta_2] + l(a_2,\zeta_2)P[a_2 \mid \zeta_2] = 40 \times 0.5 = 20$$



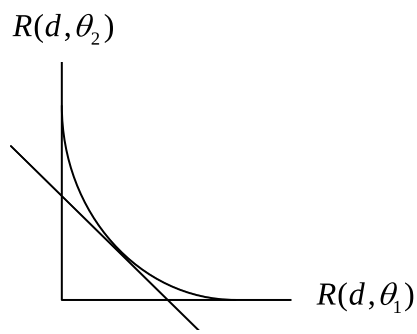

Clearly the decision strategy $d_4$ is inadmissible.

If $P[A] = 1/3,$ and $P[B] = 2/3$ we compute the average risks associated with each strategy as

$$B(d_1) = 1/3 \times 2 + 2/3 \times 20 = 14.3$$
$$B(d_2) = 1/3 \times 0 + 2/3 \times 40 = 26.7$$
$$B(d_3) = 1/3 \times 10 + 2/3 \times 0 = 3.3$$
$$B(d_4) = 1/3 \times 8 + 2/3 \times 20 = 16$$

So clearly, the Bayes strategy is $d_3$ i.e., don't buy a battery at this shop.

7.8. From the given information, the admissible strategies are on the curve shown below:




$$0.5R(d,\theta_1)+0.5R(d,\theta_2)=c*$$

The value of $c*$ is the point at which they touch. For the curve we have

$\dfrac{dR(d,\theta_2)}{dR(d,\theta_1)} = -\dfrac{[R(d,\theta_1)-1]}{[R(d,\theta_2)-1]}$ while for the line we have $\dfrac{dR(d,\theta_2)}{dR(d,\theta_1)} = -1.$

Hence at the point that they touch $-\dfrac{[R(d,\theta_1)-1]}{[R(d,\theta_2)-1]} = -1$ and $R(d,\theta_2)=R(d,\theta_1)$ . Thus the Bayes

strategy corresponds to the decision function for which $R(d,\theta_2)=R(d,\theta_1)$ .

---

.

7.9 We are given $S_1 \triangleq (-\infty,0),\ S_2 \triangleq (0,\infty);\ \mu_1 = 1/2,\ \mu_2 = -1/2$ and $X : N(\mu,1)$. Hence

$$P[X \in S_1 \,|\, 1/2] = (2\pi)\int_{-\infty}^{0} \exp\left(-0.5(x-1/2)^2\right)dx = F_{SN}(-1/2) = 0.3085$$

$$P[X \in S_2 \,|\, 1/2] = (2\pi)\int_{0}^{\infty} \exp\left(-0.5(x-1/2)^2\right)dx = F_{SN}(1/2) = 0.6915$$

$$P[X \in S_1 \,|\, -1/2] = (2\pi)\int_{-\infty}^{0} \exp\left(-0.5(x+1/2)^2\right)dx = F_{SN}(1/2) = 0.6915$$

$$P[X \in S_2 \,|\, -1/2] = (2\pi)\int_{0}^{\infty} \exp\left(-0.5(x+1/2)^2\right)dx = F_{SN}(-1/2) = 0.3085$$

The four decision functions are:

$d_1(X): d_1(X \in S_1) = a_2;\ d_1(X \in S_2) = a_1$

$d_2(X): d_2(X \in S_1) = a_1;\ d_2(X \in S_2) = a_1$

$d_3(X): d_3(X \in S_1) = a_2;\ d_3(X \in S_2) = a_2$

$d_4(X): d_4(X \in S_1) = a_1;\ d_4(X \in S_2) = a_2$

The loss functions are $l(a_1, \mu_1) = 0,\ l(a_1, \mu_2) = 2,\ l(a_2, \mu_1) = 5,\ l(a_2, \mu_2) = 0$

The risk functions are:

$$R(d_1, 1/2) = l(a_1, 1/2)P[a_1 | 1/2] + l(a_2, 1/2)P[a_2 | 1/2] = 0 + 5 \times 0.3085 = 1.54$$

$$R(d_1, -1/2) = l(a_1, -1/2)P[X \in S_2 | 1/2] + l(a_2, -1/2)P[X \in S_1 | 1/2] = 2 \times 0.3085 = 0.61$$

$$R(d_2, 1/2) = l(a_1, 1/2)P[a_1 | 1/2] + l(a_2, 1/2)P[a_2 | 1/2] = 0$$

$$R(d_2, -1/2) = l(a_1, -1/2)P[a_1 | 1/2] + l(a_2, -1/2)P[a_2 | 1/2] = 2$$

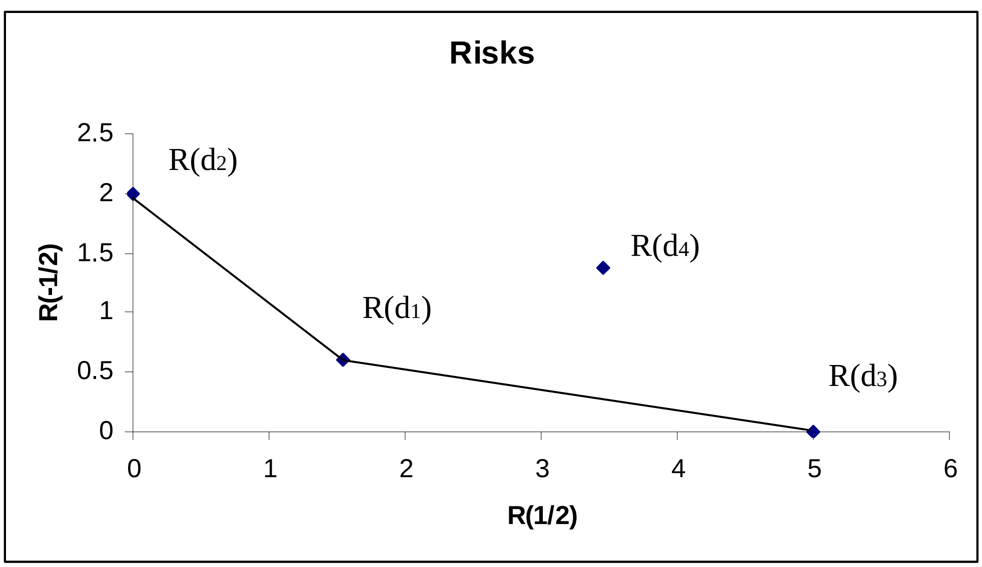
$$R(d_3, 1/2) = l(a_1, 1/2)P[a_1 | 1/2] + l(a_2, 1/2)P[a_2 | 1/2] = 0 + 5 = 5$$

$$R(d_3, -1/2) = l(a_1, -1/2)P[a_1 | 1/2] + l(a_2, -1/2)P[a_2 | 1/2] = 0$$

$$R(d_4, 1/2) = l(a_1, 1/2)P[a_1 | 1/2] + l(a_2, 1/2)P[a_2 | 1/2] = 0 + 5 \times 0.6915 = 3.46$$

$$R(d_4, -1/2) = l(a_1, -1/2)P[a_1 | -1/2] + l(a_2, -1/2)P[a_2 | 1/2] = 2 \times 0.6915 = 1.383.$$

So the four decision functions lead to conditional risks (1.54, 0.61), (0, 2), (5, 0), (3.46, 1.38).



7.10 Assume that we have *m* samples from population $X_1$ and *n* samples from population $X_2$. We form the RV

$$T = \frac{\sqrt{mn/(m+n)}\left(\hat{\mu}_1 - \hat{\mu}_2\right)}{\sqrt{\left(\sum_{i=1}^{m}(X_{1i} - \hat{\mu}_1)^2 + \sum_{i=1}^{n}(X_{2i} - \hat{\mu}_2)^2\right)/\left(m+n-2\right)}}$$

and note that $H_1$, $T$ has the t-distribution with *m*+*n*-2 degrees of freedom. The critical region is the event $T^2 > t_c^2$. To find $t_c^2$ at the 0.05 level we need to solve $P$[reject $H_1 \mid H_1$ true]=$P\ [T^2 > t_c^2] = 0.05$, or, equivalently, $P[-t_c < T < t_c] = 0.05$. Hence we find that $F_T(t_{0.025}) = 0.975$ or $t_{0.025} = x_{0.975}$, the 97.5 percentile of the $T$ RV. We find $x_{0.975}$ for a given *m*+*n*-2. For example if *m*=*n*=8, $t_{0.025} = 2.12$. Thus if $T^2 > 4.5$, the hypothesis is rejected.

---

7.11. The test for $H_1: \mu_1 = \mu_2$ versus $H_2: \mu_1 > \mu_2$ the critical region would be $T > t_{0.05} = x_{0.95}$, which we find from $F_T(x_{0.95}) = 0.95$ for the appropriate DOF.

---

7.12. The test for $H_1: \mu_1 = \mu_2$ versus $H_2$: $\mu_1 < \mu_2$ the critical region would be $T < -t_{0.05}$, which we find from $F_T(x_{0.95}) = 0.95$ for the appropriate DOF. Then $-t_{0.05} = -x_{0.95}$

---

7.13. The test for $H_1: \sigma^2 = \sigma_0^2$ versus $H_2$: $\sigma^2 \neq \sigma_0^2$ is done using the Chi-square RV as a statistic i.e. $W_{n-1} \triangleq \sum_{i=1}^{n}(X_i - \hat{\mu}_X)^2/\sigma_0^2$. The critical region is of the form $0 < W_{n-1} < a$ and $b < W_{n-1} < \infty$. An approximate solution for a,*b* is given by $F_{\chi^2}(a) = \alpha/2$ so that $a = x_{\alpha/2}$ i.e. the $x_{\alpha/2}$ percentile of $W_{n-1}$. Also $F_{\chi^2}(b) = 1-\alpha/2$ so that

$b = x_{1-\alpha/2}$.

---

7.14. a) The LRT ( for acceptance of the hypothesis) is

$$\Lambda = \frac{(2\pi\sigma^2)^{-1/2}\exp\left(-0.5(X-1)^2/\sigma^2\right)}{(2\pi\sigma^2)^{-1/2}\exp\left(-0.5X^2/\sigma^2\right)} = \exp\left((X-1/2)/\sigma^2\right) > k .$$

b) take natural logs of both sides, obtain $X > \sigma^2 \ln k + 1/2 = c$ where $c \triangleq k\sigma^2 + 1/2$.

c) $\alpha = 0.02 = \int_{\infty}^{c}\left(2\pi\sigma^2\right)^{-1/2}\exp\left(-0.5(x-1)^2/\sigma^2\right)dx = F_{SN}(z_{0.02}) = F_{SN}\left((c-1)/\sigma\right)$.

Hence $c = z_{0.02}\sigma + 1$.

d) From the tables we find $z_{0.02} = -2.05$, hence with $\sigma = 1$ we find $c = -1.05$

---

7.15 The LRT for acceptance of $H_1$ is

$\Lambda = \dfrac{(2\pi)^{-1/2}\exp\left(-0.5\sum_{i=1}^{n}(X_i-3)^2\right)}{(2\pi)^{-1/2}\exp\left(\left(-0.5\sum_{i=1}^{n}(X_i-1)^2\right)\right)} > k$. Simplifying, we get accept $H_1$ if $4n\hat{\mu} > \ln k + 8n$, or

if $\hat{\mu} > \dfrac{1}{4n}\ln k + 2 \triangleq c_n$. To find $c_n$, we solve

$\alpha = P[\text{reject } H_1 \mid H_1 \text{ true}] = F_{SN}\left(\sqrt{n}(c_n - 3)\right) = F_{SN}(z_\alpha)$. Thus the main result is

$\sqrt{n}(c_n - 3) = z_\alpha$ or $c_n = \dfrac{z_\alpha}{\sqrt{n}} + 3$. For $\alpha = 0.01, z_\alpha = -2.33$ and $c_n$ =2.26.

---

7.16 We are given $\alpha = 0.02,\ \beta = 0.01$. Hence the power of the test is

$1-\beta = 0.99 = \left(2\pi n^{-1}\right)^{-1/2}\int_{-\infty}^{c_n}\exp\left(-0.5n(x-1)^2\right)dx = F_{SN}\left(\sqrt{n}(c_n-1)\right) = F_{SN}(z_{0.99})$, while from Problem 7.15 we have

$\alpha = P[\text{reject } H_1 \mid H_1 \text{ true}] = F_{SN}\left(\sqrt{n}(c_n - 3)\right) = F_{SN}(z_\alpha)$.

Hence we have two equations in two unknowns:

$\sqrt{n}(c_n - 1) = z_{0.99} = 2.33$ and $\sqrt{n}(c_n - 3) = z_{0.02} = -2.05$ .

Solving we get $c_n = 2.2,\ n = 5$ (rounded up from 4.5).

---

7.17. To keep at $\alpha = 0.02,$ we have to satisfy $c_n = \frac{z_{0.02}}{\sqrt{n}} + 3 = \frac{-2.05}{\sqrt{n}} + 3$. Next, to satisfy a level of power $\mathrm{P}=1-\beta$ we have to satisfy $\sqrt{n}(c_n - 1) = z_{\mathrm{P}}$. Substituting for $c_n$ and simplifying yields $n = \left(0.5(z_{\mathrm{P}} + 2.05)\right)^2$. Now go to Excel and create three columns: P, Norm(P,0,1), and *n*. Noted that Norm(P,0,1) returns $z_{\mathrm{P}}$ from which we can compute *n*. Finally call for the Chart Wizard to produce the required graph.

---

7.18 **m.file:** The following m.file computes realizations of the unbiased sample variances of the two populations P1 and P2.. In this case P1 and P2 are $N(0,1)$

function [sigs1,sigs2,lamb]=ftest(n1,n2)

y1=normrnd(0,1,[1,n1]); % this is the Matlab call to generate n1 $N(0,1)$ realizations from P1.

y2=normrnd(0,1,[1,n2]); ]); % this is the Matlab call to generate n2 $N(0,1)$ realizations from P2 .

mu1=sum(y1)/n1; % computes the sample mean of P1

mu2=sum(y2)/n2; % computes the sample mean of P2

z1=(y1-mu1).^2; % this is unnecessary but shows how each point in the P1 data array is squared.

z2=(y2-mu2).^2; % this is unnecessary but shows how each point in the P2 data array is squared.

sigs2=sum((y2-mu2).^2)/(n2-1);

sigs1=sum((y1-mu1).^2)/(n1-1)

**Command widow:**

>> [sigs1,sigs2]=ftest(11,21) % this commands calls for 11 samples from P1 and 21 samples from P2

sigs1 =

0.9855

sigs2 =

0.9823

>> V=sigs1/sigs2

V =

1.0033

**F-test:** at the $\alpha = 0.05$ level we find $F_F(x_{0.025};10,20) = 0.025 \rightarrow x_{0.025} = 0.36$ and similarly $F_F(x_{0.975};10,20) = 0.025 \rightarrow x_{0.975} = 2.77$. Since 0.025<V<2.77. we accept the hypothesis that the two populations have the same variance.

---

7.19 In this problem we have *k* groups (sub-clusters) and the *i*th group $(i = 1,...,k)$ is populated with $n_i$ (assumed continuous) i.i.d. observations $Y_{ij},\ j = 1,...,n_i$. With $Z_i = (1/n_i)\sum_{j=1}^{n_i} Y_{ij} = \hat{\mu}_{Y_i}$ we find that $\mu_{Z_i} = \mu_{Y_i}$ and $\sigma_{Z_i}^2 = \sigma_{Y_i}^2 / n_i$ and if $n_i >> 1$, (assumed true for each $i$), then $Z_i \xrightarrow{n_i \to \infty} N(\mu_{Z_i}, \sigma_{Z_i}^2)$. The quantity $\sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i)^2$ can be regarded as the *within-group variability* of the ith sub-cluster. The sequence of RVs $Z_i,\ i = 1,...,k$ can then be taken as Normal RVs with (typically) different means and variances. However the sequence of RVs $V_i \triangleq (Z_i - \mu_{Z_i})/\sigma_{Z_i},\ i = 1,...,k$ are $N(0,1)$. Hence $\sum_{i=1}^{k} V_i^2 : \chi_k^2$. Also with $W_i \triangleq (Z_i - \hat{\mu}_{Z_i})/\sigma_{Z_i},\ i = 1,...,k$ we find that $\sum_{i=1}^{k} W_i^2 : \chi_{k-1}^2$, one degree-of-freedom being lost in replacing the mean by the sample mean. The quantity $\hat{\mu}_Z \triangleq (1/k)\sum_{i=1}^{k} Z_i$ is the overall sample mean of the *k* RVs $Z_i, i = 1,...,k$. It represents the "center of gravity" RV of the cluster $\{Z_i,\ i = 1,...,k\}$ and hence of all the data. Note that $\mu_Z \triangleq (1/k)\sum_{i=1}^{k} \mu_{Z_i}$. It is not unreasonable to regard $\sum_{i=1}^{k} (Z_i - \hat{\mu}_Z)^2$ as the RV that measures *inter-group variability* since it measures the total distance-squared from the center-of-gravity of each sub-cluster $\{Y_{ij},\ j = 1,...,n_i\}$ to $\mu_Z$.

---

7.20 We assume that the within-cluster data $\{Y_{ij},\ j = 1,...,n_i\}$ are i.i.d for each value of *i*. We assume also that between-cluster data are independent. Thus all the data is independent, whether in a specific cluster or not. Now consider the double sum

$S \triangleq \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left( \frac{Y_{ij} - Z_i}{\sigma_{Y_i}} \right)^2$. The inner sum is $\chi_{n_i - 1}^2$ so that *S* is the sum of *k* independent *Chi-*square RVs, written, somewhat ingloriously, as $S \triangleq \chi_{n_1 - 1}^2 + \chi_{n_2 - 1}^2 + \cdots + \chi_{n_k - 1}^2$. The MGF of a Chi-square RV with *m* degrees of freedom is $(1-2t)^{-m/2}$ for $t<1/2$. The MGF of the sum of *k* i.i.d Chi-square RVs is the *k*-product of their MGFs, which in this case is

$(1-2t)^{-(\sum_{i=1}^{k} n_i - k)/2}$ for $t<1/2$. Hence *S* is Chi-square as $\chi^2_{n-k}$ where we recalled that $\sum_{i=1}^{k} n_i = n$.

To test for the hypothesis that all the groups are the same i.e. all the data in all the groups come from the same population we construct an F-statistic as follows: *we assume that the* $Z_i$ *are i.i.d.* i.e. $E[Z_i]=\mu_{Y_i} \triangleq \mu_Y$ with $Var[Z_i]=\sigma^2_{Y_i}/n_i=\sigma^2_Y/n_i$. Then $Z_i : N(\mu_Y,\sigma^2_Y/n_i)$ and

$\sum_{i=1}^{k} n_i \left((Z_i-\hat{\mu}_Z)/\sigma_Y\right)^2$ is $\chi^2_{k\text{-}1}$ . Then the ratio

$$F_{k-1,n-k}=\left(\sum_{i=1}^{k} n_i \left(Z_i-\hat{\mu}_Z\right)^2/(k-1)\right)/\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij}-Z_i\right)^2/(n-k)\right)$$

is the appropriate statistic is for testing the hypothesis that all the data come from the same population.

7.21 We use the F-test statistic

$$F_{k-1,n-k}=(n-k)\sum_{i=1}^{k} n_i \left(Z_i-\hat{\mu}_Z\right)^2/(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij}-Z_i\right)^2,$$

which we rewrite for convenience as

$$F_{k-1,n-k}=\left(\sum_{i=1}^{k} n_i \left(Z_i-\hat{\mu}_Z\right)^2/(k-1)\right)/\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij}-Z_i\right)^2/(n-k)\right)$$

and identify $n=1000,\ n_i=200,\ i=1,...,5$ and $k=5$. From the data we are given

$Z_1=3.17, Z_2=2.72, Z_3=2.63, Z_4=2.29, Z_5=2.19$ yielding $\hat{\mu}_Z=2.6$.Then the numerator is computed as

$200\times\left((3.17-2.6)^2+(2.72-2.6)^2+(2.63-2.6)^2+(2.29-2.6)^2+(2.19-2.6)^2\right)/4=30.225$. To compute denominator, we have to assume that the standard deviations were computed as

$\sigma_i=\left((1/199)\times\sum_{j=1}^{200}(Y_{ij}-Z_i)^2\right)^{1/2},\ i=1,...,5$. Then it follows that $199\sigma_i^2=\sum_{j=1}^{200}(Y_{ij}-Z_i)^2$, and the denominator is computed as $199\times\left(0.74^2+0.71^2+0.73^2+0.70^2+0.72^2\right)/795\text{=}0.65$. Hence the F-statistic is computed as 30.225/0.65=46.5.

Next we go to the F-test calculator e.g. BioKin on line and enter the degrees of freedom $\nu_1 = 4$ (numerator) $\nu_2 = 495$ (denominator) and the significance level 0.05 for a one-sided test and obtain 2.37. Since 46.5>> 2.37 the hypothesis is strongly rejected.

---

7.22 .We use Pearson's Chi-square test as follows. The expected number of green seeds is $880 \times 0.75 = 660$ while the expected number of yellow seeds is $880 \times 0.25 = 220$. Pearson's statistic yields

$$\Lambda' = \frac{(660-639)^2}{880 \times 0.75} + \frac{(241-220)^2}{880 \times 0.25} = 0.67 + 2.0 = 2.67$$

At the 0.05 level of significance the hypothesis is accepted if $0.001 < \Lambda < 5.02$. Hence the hypothesis is accepted.

---

7.23. (t-test)

We are given two sets of realizations and told that they come from Normal distributions with the same variance. We use the t-test to test $H_1 : \mu_1 = \mu_2$ versus $H_2 : \mu_1 \neq \mu_2$

**Set 1:**
-5.980e-1 -9.290e-1 -8.340e-2 1.020e+0 6.780e-1 2.890e-1 1.430e-1 -2.060e+0 1.260e+0 1.670e+0

**Set 2:**
6.270e-1 2.640e+0 1.530e+0 5.920e-1 1.910e+0 5.050e-1 7.660e-1 2.760e-1 3.070e+0 8.550e-1

Using Excel we compute $\hat{\mu}_1 = 0.14$, $\hat{\mu}_2 = 1.28$. Also writing the t-statistic as *T=NUM/DEN* where $NUM = (\hat{\mu}_1 - \hat{\mu}_2) \times \sqrt{nm/(m+n)}$ and

$DEN = \left( \sum_{i=1}^{m} (X_{1i} - \hat{\mu}_1)^2 + \sum_{j=1}^{n} (X_{2j} - \hat{\mu}_2)^2 \right)^{1/2} /(m+n-2)^{1/2}$ we obtain

$T = NUM / DEN = -2.54/1.4 = -1.82$ and $T^2 = 3.31$.This is a two-sided test with 0.025 error probability assigned to each tail. Thus if $T < -t_{0.025}$ or $T > t_{0.025}$ where

$\int_{-\infty}^{t_{\alpha/2}} f_T(x; m+n-2)dx = 1-\alpha/2$. From this we get that $t_{0.025} = 2.1$ At a overall significance of 0.05, the hypothesis is rejected if the t-statistic lies outside the interval (-2.1, 2.1). Since -1.82 is inside this interval, the hypothesis $H_1$ is accepted.

---

7.24 Under $H_1$ we have $p_i = p_{0i}$ all $i$; hence

$$E[V \mid H_1] = E\left(\sum_{i=1}^{l}(np_{0i})^{-1}(n_i^2 - 2nn_i p_{0i} + n^2 p_i^2)\right)$$
$$= \sum_{i=1}^{l}(np_{0i})^{-1} E(n_i^2 - 2nn_i p_{0i} + n^2 p_i^2)$$
$$= \sum_{i=1}^{l}(np_{0i})^{-1}(np_{0i}(1-p_{0i}) + n^2 p_{0i}^2 - 2n^2 p^2{}_{0i} + n^2 p_{0i}^2)$$
$$= \sum_{i=1}^{l}(np_{0i})^{-1}(np_{0i}(1-p_{0i}) = l-1$$

7.24 **Solution:** Let us compute $\Lambda$ under the assumption that $H_1$ is true. Using that $\sum_{i=1}^{m}\left(X_{1i}-\hat{\mu}_1\right)^2 = \sigma^2\chi_{m-1}^2$, $(W_{m-1} : \chi_{m-1}^2)$and $\sum_{i=1}^{n}\left(X_{2i}-\hat{\mu}_2\right)^2 = \sigma^2\chi_{n-1}^2$ $(W_{n-1} : \chi_{n-1}^2)$ ~~,~~ and factoring out constants we get that

$$\Lambda = A(m,n)\frac{\left(\dfrac{1}{\chi_{m-1}^2 + \chi_{n-1}^2}\right)^{(m+n)/2}}{\left(\dfrac{1}{\chi_{m-1}^2}\right)^{m/2}\left(\dfrac{1}{\chi_{n-1}^2}\right)^{n/2}}, \qquad (7\cancel{6}.3\text{-}16)$$

where. We recall that the random variable $F_{m,n}$ defined as

$F_{m,n} \triangleq \dfrac{\chi_m^2/m}{\chi_n^2/n}$ is said to have the F-distribution with $m$ and $n$ degrees of freedom respectively (the numerator DOF is cited first). Then rewriting $\Lambda$ as

$$\Lambda = A(m,n)\frac{\left(\dfrac{n-1}{(n-1)\times\chi_{n-1}^2}\right)^{m/2}\left(\dfrac{1}{\chi_{n-1}^2}\right)^{n/2}\left(\dfrac{1}{[(m-1)/(n-1)\times((n-1)\chi_{m-1}^2/(m-1)\chi_{n-1}^2)]+1}\right)^{(m+n)/2}}{\left(\dfrac{m-1}{(m-1)\times\chi_{m-1}^2}\right)^{m/2}\left(\dfrac{1}{\chi_{n-1}^2}\right)^{n/2}}$$

it follows that

$$\Lambda = A(n,m)\frac{\left(\frac{(n-1)}{(m-1)}F_{n-1,m-1}\right)^{n/2}}{\left(1+\frac{(n-1)}{(m-1)}F_{n-1,m-1}\right)^{(m+n)/2}}$$

$$= A(m,n)\frac{\left(\frac{(m-1)}{(n-1)}F_{m-1,n-1}\right)^{m/2}}{\left(1+\frac{(m-1)}{(n-1)}F_{m-1,n-1}\right)^{(m+n)/2}}.$$

---

7.25 Under $H_2$ we have $p_i = p_{1i}$ all $i$ ; hence

$$E[V \mid H_2] = E\left(\sum_{i=1}^{l}(np_{0i})^{-1}(n_i^2 - 2nn_i p_{0i} + n^2 p_{0i}^2)\right)$$

$$= \sum_{i=1}^{l}(np_{0i})^{-1}E(n_i^2 - 2nn_i p_{0i} + n^2 p_{0i}^2)$$

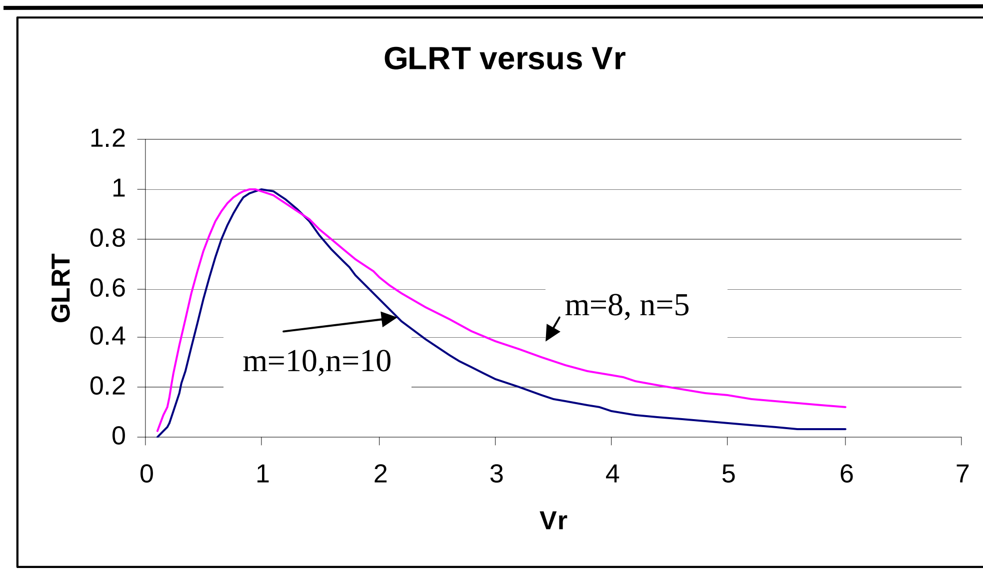
$$= \sum_{i=1}^{l}(np_{0i})^{-1}(np_{1i}(1-p_{1i}) + n^2 p_{1i}^2 - 2n^2 p_{0i}p_{1i} + n^2 p_{0i}^2)$$

$$= \sum_{i=1}^{l}(np_{0i})^{-1}\left((np_{1i}(1-p_{1i}) + n^2(p_{0i}-p_{1i})^2\right).$$

If we differentiate with respect each of the $p_{0i}$ and set the result equal to zero, we find that the minimum occurs when $p_{0i} = p_{1i}$ all $i$ . However, under $H_2$ this cannot be since at least two of the $p_{0i} \neq p_{1i}$ . Clearly $E[V \mid H_2]$ becomes unbounded for $p_{0i}$ close to zero and/or when $n \to \infty$.

7.26 The plot is produced using Excel.

---



At the 0.05 level this is a two sided test on the RV $V_R$ . We seek

the percentile solutions to $F_T(x_{0.025}) = 0.025$ and $F(x_{0.975}) = 0.975$.

The solutions are $x_{0.025} = 0.18$ and $x_{0.975} = 9.07$. The test is: reject

$H_1$ if $V_r < x_{0.025} = 0.18$ or $V_R > x_{0.975} = 9.07$

---

7.27. Most of the proof is given in the text. However, we note that
in the text we introduce

$$(m-1)\hat{\sigma}_1^2 = \sum_{i=1}^{m}(X_{1i} - \hat{\mu}_1)^2$$
$$(n-1)\hat{\sigma}_2^2 = \sum_{j=1}^{n}(X_{2j} - \hat{\mu}_2)^2$$

and under $H_1$, $\sigma_1^2 = \sigma_2^2 = \sigma^2$ so that

$$\begin{aligned}(m-1)\hat{\sigma}_1^2 / \sigma^2 &= \sum_{i=1}^{m}\left((X_{1i} - \hat{\mu}_1)/\sigma\right)^2 \triangleq W_{m-1} : \chi^2_{m-1} \\ (n-1)\hat{\sigma}_2^2 / \sigma^2 &= \sum_{i=1}^{n}\left((X_{2i} - \hat{\mu}_2)/\sigma\right)^2 \triangleq W_{n-1} : \chi^2_{n-1}\end{aligned}.$$

The variable $F \triangleq \dfrac{(m-1)^{-1}W_{m-1}}{(n-1)W_{n-1}}$ has the F-distribution. Hence $\hat{\sigma}_1^2 / \hat{\sigma}_2^2 : F_{m-1,n-1}$.

---

7.28 The theory behind the test is as follows. The likelihood function is

$L(\mu, \sigma^2) = (2\pi\sigma^2)^{n/2} \exp\left(-0.5\left(\sum_{i=1}^{n}[(X_i - \mu)/\sigma]^2\right)\right)$. The global i.e. unrestricted maximum is

obtained by differentiating with respect to $\mu$ and $\sigma^2$. The differentiation

yields $\mu^\dagger = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \hat{\mu},\ \sigma^{2\dagger} = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$ .

Next, finding the (local) maximum under $H_1$ yields

$\mu^* = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \hat{\mu},\ \sigma^{2*} = \sigma_0^2$. Taking the ration of $L(\hat{\mu}, \sigma_0^2) / L(\hat{\mu}, \sigma^{2\dagger}) \triangleq \Lambda$ yields

$\Lambda = (W/n)^{n/2} \exp\left(-0.5(W-n)\right)$, where $W$ under $H_1$ is Chi-square with DOF $n$-1. Specifically

$W = \sum_{i=1}^{n} [(X_i - \hat{\mu})/\sigma_0]^2$. A plot of $\Lambda$ versus $W$ has the appearance as in Figure 7.3-8. Hence acceptance of $H_1$ requires that $\Lambda > \lambda_C$ or, equivalently, that $w_l < W < w_u$ ,where $w_l, w_u$ are determined from the type I error criterion and the "equal error rule" discussed in the text. Thus given that $\alpha = P[\text{reject } H_1 \mid H_1 \text{ is true}]$. Thus we seek $\alpha/2 = F_{\chi^2}(w_l; n-1)$ and $1-\alpha/2 = F_{\chi^2}(w_u; n-1)$ . Thus we recognize that $w_l = x_{\alpha/2}$ i.e. the $\alpha/2$ percentile point and $w_u = x_{1-\alpha/2}$ i.e. the $1-\alpha/2$ percentile point.

Summary for testing $H_1 : \sigma^2 = \sigma_0^2$ versus $H_2 : \sigma^2 \neq \sigma_0^2$:

1. Obtain realizations $x_1, x_2, ..., x_n$ of $X_1, X_2, ..., X_n$ respectively;

2. Compute the realization of $W$ as $w = \sum_{i=1}^{n} [(x_i - \hat{\mu}')/\sigma_0]^2$, where $\hat{\mu}' = \frac{1}{n}\sum_{i=1}^{n} x_i$ is a realization of $\hat{\mu}$

2. Choose the significance level of the test $\alpha$ e.g. $0.1,\ 0.05,\ 0.025,\ 0.01$;

3. From the tables of the Chi-square CDF find the values $w_l = x_{\alpha/2}$ and $w_u = x_{1-\alpha/2}$ for $n$-1;

4. If $w_l < w < w_u$, accept $H_1$, else reject it.

---

7.29 We use the Pearson test on $H_1:\ P[\text{Heads}]=0.5$ versus $P[\text{Heads}]>0.5$. Hence $V = \left(\frac{35-50\times 1/2}{\sqrt{50\times 1/2}}\right)^2 + \left(\frac{15-50\times 1/2}{\sqrt{50\times 1/2}}\right)^2 = 8$. At the 0.05 level of significance we find $F_{\chi^2}(0.95;1) = 3.84$. Since 8>3.84 we reject $H_1:\ P[\text{Heads}]=0.5$ .

---

7.31 We estimate the 100ith percentile from $\frac{100i}{(n+1)}$. In this case $n$=24. We note that $y_7 \sim x_{0.28}$, $y_8 \sim x_{0.32}$. Hence $x_{0.3} \sim y_2 + \frac{0.3-0.28}{1/25}(y_3 - y_2) = (y_2 + y_3)/2$. Thus $(y_2 + y_3)/2$ estimates the 30th percentile point.

---

7.32. We compute for ordered co-joined sequence *d*=14. From Example 7~~6~~.5-8, the critical value is $d_0 = 6.3$. Since $d_0 < d$ we accept the hypothesis that $\mathrm{P}_1 = \mathrm{P}_2$. Yet it is obvious that $\mathrm{P}_1$ generates even numbers while $\mathrm{P}_2$ generates odd numbers. The run test is not sensitive to populations with all even/odd parity.

---

7.33.It is clear from the data that if the population generating the $S_1$ data is $X$ and the population generating the $S_2$ data is $Y$ then $Y = -10 \times X$ . So the correlation coefficient is -1. So in this sense the source are the same since given *Y* you can get *X*. However since *d*=2, the run test result would say that the sources are different.

---

7.34 We compute for ordered co-joined sequence *d*=14. From Example 7~~6~~.5-8, the critical value is $d_0 = 6.3$. Since $d_0 < d$ we accept the hypothesis that $\mathrm{P}_1 = \mathrm{P}_2$. Yet it is obvious that $\mathrm{P}_1$ generates even numbers while $\mathrm{P}_2$ generates odd numbers. The run test is not sensitive to populations with all even/odd parity.