Homework Nine

Jeremy Harper
STA4702
03.14.12

5.18a. As the *n* is large, we can use the $\chi^2_{(p)}$ critical value, and reject the null hypothesis if $T^2 > \chi^2_{(p)}(\alpha)$ (Result 5.4). For $\chi^2_{(3)}(0.05)$, the critical value is 7.814. Since the T^2 is larger than $\chi^2_{(3)}(0.001)$ (223.2 > 7.814), we can reject the null hypothesis that the current scores are not different from the past 10 year scores at the 0.05 level.

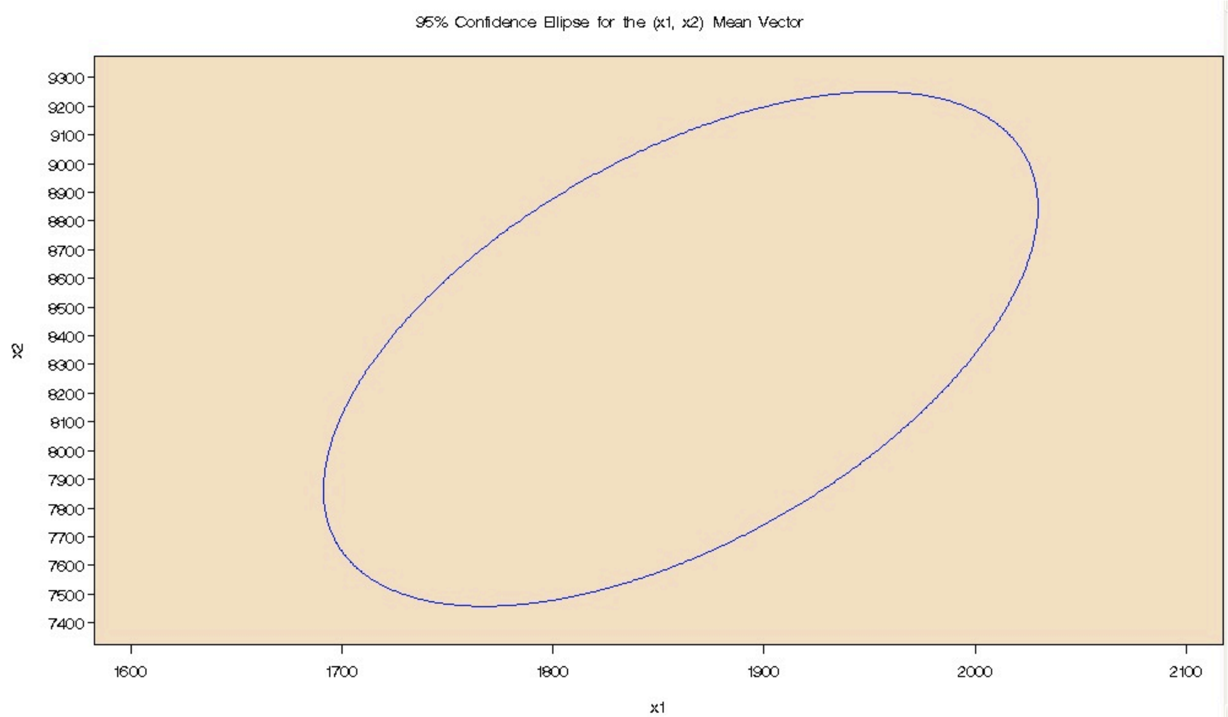
To use the normal theory test (Statement 5-7) where T^2 is distributed as $((n-1)p/(n-p))*F_{p,\ n-p}$, and $F_{3,\ 87-3}(0.05) = 2.27$, the critical value for rejecting the null hypothesis is 6.97. Based on Hotelling's T^2, we can conclude that the current scores are statistically different than those over the last 10 years (reject null hypothesis), since $T^2 > F$ (223.3 > 6.97).

mu0	xbar
500	526.58621
50	54.689655
30	25.126437

s			sinv		
5808.0593	597.8352	222.02967	0.0004319	-0.001574	-0.002557
597.8352	126.05373	23.388532	-0.001574	0.0155044	-0.000567
222.02967	23.388532	23.111735	-0.002557	-0.000567	0.0684014

t2	f	df1	df2	p
223.31018	72.705639	3	84	<0.0001

5.19a.



Half-length of major and minor axes:

x1. 901.65222
x2. 140.51187

```
 Variable n    xbar    variance     t1

    x1    30  1860.50   124054.67  2.04523
    x2    30  8354.13  3486333.15  2.04523

      f        losimultaneousCI     upsimultaneousCI

    3.34039         1691.35             2029.65
    3.34039         7457.41             9250.85

                            S

                124054.67 361620.45
                361620.45 3486333.2

            EvecS            EvalS            F

    0.1057399 0.9943938   3524786.5      3.3403856
    0.9943938  -0.10574   0 85601.376
```

```
/* 18a */

data t52;
  infile "\\psf\Home\Documents\University\Spring_2012\STA4702\Datasets\T5-2.dat";
  input x1 x2 x3;
run;

proc iml;
  start hotel;
```

```
    mu0={500, 50, 30};
    one=j(nrow(x),1,1);
    ident=i(nrow(x));
    ybar=x`*one/nrow(x);
    s=x`*(ident-one*one`/nrow(x))*x/(nrow(x)-1.0);
     sinv = inv(s);
    print mu0 ybar;
    print s sinv;
    t2=nrow(x)*(ybar-mu0)`*inv(s)*(ybar-mu0);
    f=(nrow(x)-ncol(x))*t2/ncol(x)/(nrow(x)-1);
    df1=ncol(x);
    df2=nrow(x)-ncol(x);
    p=1-probf(f,df1,df2);
    print t2 f df1 df2 p;
  finish;
  use t52;
  read all var{x1 x2 x3} into x;
  print x;
  run hotel;
quit;



/* 19a */

data t511;
  infile "\\psf\Home\Documents\University\Spring_2012\STA4702\Datasets\T5-11.dat";
  input x1 x2;
run;

/* Program to graph the confidence ellipse for the mean vector */
%let inputdata = t511;  /* this line must be edited */
%let var1      = x1  ;           /* this line must be edited */
%let var2      = x2    ;           /* this line must be edited */
%let conf      = 95    ;           /* Confidence level desired*/
proc corr data=&inputdata noprint nocorr cov outp=covout(type=cov);
    var &var1 &var2;
run;
data covonly;
   set covout;
   if _type_='COV';
   keep &var1 &var2;
run;
data meanonly;
   set covout;
   if _type_='MEAN';
   keep &var1 &var2;
run;
data nonly;
   set covout;
   if _type_='N';
   keep &var1 &var2;
run;
proc iml;
   use covonly;
   read all into S;
   p = ncol(S);
   use meanonly;
   read all into xbar;
   xbar = xbar`;
   use nonly;
   read all into n;  n=n[1,1];
  A = S/n;
  Evec = Eigvec(A);
  Eval = diag(Eigval(A));
  EvecS = Eigvec(S);
  EvalS = diag(Eigval(S));
  try1 = Evec*Eval*Evec`;
  center = xbar;
  F = finv(&conf/100, p, n-p);
```

```
  one = (p*(n-1));
  two = (n*(n-p));
  diff = ((p*(n-1))/(n*(n-p)))*F;
  diffT = sqrt(((p*(n-1))/(n-p))*finv(&conf/100, p, n-p));
  ss = sqrt(diag(S)/n);
  distances = sqrt(diag(EvalS))*sqrt(((p*(n-1))/(n*(n-p)))*finv(&conf/100, p, n-p));
  distance = sqrt((n-1)*p*finv(&conf/100, p, n-p) /(n-p));
  T2distminus   = xbar - sqrt(((p*(n-1))/(n-p))*finv(&conf/100, p, n-p))*sqrt(diag(S)/
n);
  T2distplus   = xbar + sqrt(((p*(n-1))/(n-p))*finv(&conf/100, p, n-p))*sqrt(diag(S)/
n);
  print xbar distance distances center EvecS EvalS F one two diff diffT S T2distminus
T2distplus ss;


npoints = 1000;
free xbig;
do r = 1 to npoints;
    angle = 2*3.14159265 * (r/npoints);
    w1 = sin(angle);
    w2 = cos(angle);
    w = w1//w2;
    x = Evec*sqrt(Eval)*distance*w + center;
    xbig = xbig//x`;
end;
create plotdata from xbig;
append from xbig;
quit;
goptions ftext=SWISS ctext=BLACK htext=1 cells;
axis1 width=1 offset=(3 pct) label=(a=90 r=0) ;
axis2 width=1 offset=(3 pct) ;
symbol1 c=BLUE ci=BLUE v=none height=1 cells
        interpol=spline l=1 w=1;
proc gplot data=Work.Plotdata(rename=(Col1=&var1 Col2=&var2)) ;
   title "&conf% Confidence Ellipse for the (&var1, &var2) Mean Vector";
   plot &var2 * &var1  /
      caxis = BLACK
      ctext = BLACK
      cframe = CXF7E1C2
      href=0
      vref=0
      hminor = 0
      vminor = 0
      vaxis = axis1
      haxis = axis2
      ;
      run;
quit;
goptions ftext= ctext= htext=;
symbol1;
axis1; axis2;
title;




options ls=78;
title "Confidence Intervals - t5-11";
%let p=2;
data t511_CI;
  set t511;
  variable="x1"; x=x1; output;
  variable="x2"; x=x2;    output;
  keep variable x;
  run;
proc sort;
  by variable;
  run;
proc means noprint;
  by variable;
```

```
  var x;
  output out=a n=n mean=xbar var=s2;
  run;
data b;
  set a;
  t1=tinv(1-0.025,n-1);
  tb=tinv(1-0.025/&p,n-1);
  f=finv(0.95,&p,n-&p);
  loone=xbar-t1*sqrt(s2/n);
  upone=xbar+t1*sqrt(s2/n);
  losim=xbar-sqrt(&p*(n-1)*f*s2/(n-&p)/n);
  upsim=xbar+sqrt(&p*(n-1)*f*s2/(n-&p)/n);
  lobon=xbar-tb*sqrt(s2/n);
  upbon=xbar+tb*sqrt(s2/n);
  run;
proc print;
  run;
```