STAT505 Assessment #5

1. J&W Exercise 6.19, parts a and c. The file "milk.dat" contains the data with columns fuel, repair, capital, and truck, respectively.

   (a) The null hypothesis is $H_0 : \boldsymbol{\mu_1} = \boldsymbol{\mu_2}$, where $\boldsymbol{\mu_1}$ is the mean cost vector for gasoline trucks, and $\boldsymbol{\mu_2}$ is the mean cost vector for diesel trucks. The alternative hypothesis, H_a, is that these vectors differ in at least one of their three components: fuel, repair, and capital.

   Hotelling's T^2 is 50.91, which corresponds to an F test statistic of 16.38 with 3 and 55 degrees of freedom. Since $16.38 > 4.16 = f_{3,55}(.01)$, ($p$-value ≈ 0), we can reject H_0 with significant evidence that at least one of the mean cost components differs between gasoline and diesel trucks.

   (c) Using a Bonferroni correction for multiplicity, we can say with 99% simultaneous confidence that the difference in mean fuel cost between gasoline and diesel trucks is within -1.14 to 5.37, the difference in mean repair cost is within -6.38 to 1.08, and the difference in mean capital cost is within -12.79 to -4.36.

2. The Swiss government believes that the printer is not meeting their specifications for the 1000 franc notes. The following table gives the government specifications for six different dimensions of the notes:

Dimension	Specification
Length	215 mm
Left Width	130 mm
Right Width	130 mm
Bottom Margin	9 mm
Top Margin	10 mm
Diagonal Length of Printed Area	141.354 mm

   Measurements were obtained on the above variables for a random sample of 100 notes from the printer. These data are stored in the "swiss.dat" data set.

   (a) Use Hotelling's T-squared statistic to test the null hypothesis that the printer meets government specifications on average. Give the values for the statistic, critical value, degrees of freedom, and p-value.

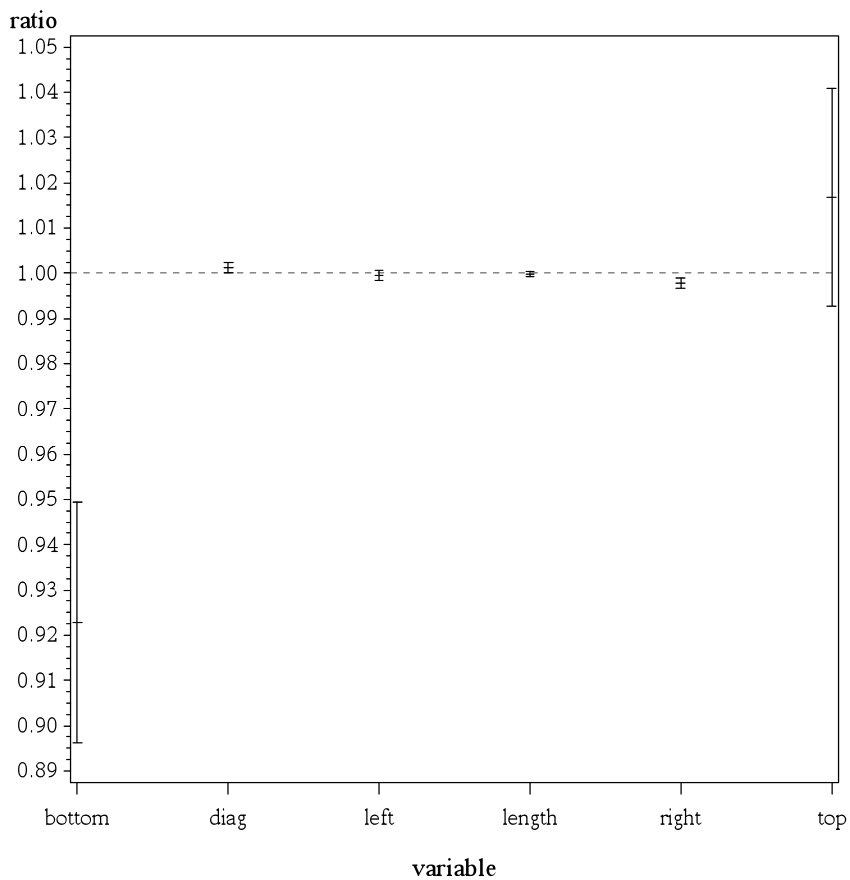
   The null hypothesis is $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$, where $\boldsymbol{\mu_0}' = [215, 130, 130, 9, 10, 141.354]$. Hotelling's value is 231.83, which exceeds $13.88 = (99(6)/94 f_{6,94}(.05)$. So, we have significant evidence that the printer does not conform to government specifications on at least one of these variables.

   (b) Find simultaneous 95% confidence intervals for the population means of the six variables. Are these intervals consistent with your conclusion in part a) above? Explain why or why not.

   Using the Bonferroni adjustment for multiplicity, the lower and upper endpoints of the simultaneous confidence intervals are given below, along with the government specified values for reference. We can see that these government values for right, bottom, and diagonal measurements do not fall within their interval estimates, which is consistent with our rejection above that all variables meet their government specified requirements.

varnames	xbar length	mu0
length	214.969	215
left	129.943	130
right	129.72	130
bottom	8.305	9
top	10.168	10
diag	141.517	141.354

(c) Give a profile plot for the ratios of the dimensions over their specifications.



(d) What can you conclude from these results? Is the printer meeting the government specifications?

Overall, we have significant evidence that the printer is *not* meeting government specifications, but the discrepancies are due to only half of the variables measured. The other half do not significantly differ from the government specifications.

3. At the start of a study to determine whether exercise or dietary supplements would slow bone loss in older women, an investigator measured the mineral content of bones. Measurements were recorded on 25 women for three bones on the dominant and non-dominant sides. The data is given in "mineral.dat". The column headings, in sequence, are dominant radius, radius, dominant humerus, humerus, dominant ulna, and ulna.

   (a) Test the equality of mean mineral contents between the dominant and non-dominant bones at the $\alpha = 5\%$ level of significance.

   The null hypothesis is $H_0 : \mu_{\mathbf{d}} = \mathbf{0}$, where μ_d is the vector of mean differences (dominant minus non-dominant) for the three different bones. The Hotelling value is 5.94, which corresponds to an F statistic of 1.82 with 3 and 22 degrees of freedom. Since the p-value is 0.17, we do not have significant evidence to say that the mean mineral values differ between dominant and non-dominant bones.

   (b) Construct 95% simultaneous confidence intervals for the mean differences.

   The Bonferroni-adjusted intervals are provided below. With 95% simultaneous confidence, the mean difference (dominant minus non-dominant) in mineral content for each bone is within its respective interval.

varnames	lobon	upbon
radius	-0.005671	0.056631
humerus	-0.007855	0.1235345
ulna	-0.029395	0.0505145

   (c) Are your answers from the two previous parts consistent with each other? What can you conclude about the dominant bones compared with the non-dominant bones?

   Considering that zero lies within each interval, it is not surprising that we failed reject the zero vector as a possibility for the mean difference vector between dominant and non-dominant mineral contents.

SAS code for this problem is below. Note that after computing the differences, the analysis proceeds identically to that for Problem 2.

```
data mineral;
infile "v:\505\datasets\mineral.dat";
input drad rad dhum hum dulna ulna;
d1=drad-rad;
d2=dhum-hum;
d3=dulna-ulna;
run;
```

SAS code for Problem 1(J&W 6.19)

```
data milk;
infile 'v:\505\datasets\milk.dat';
input fuel repair capital truck $;
run;
proc iml;
  start hotel2;
    n1=nrow(x1); n2=nrow(x2); k=ncol(x1);
    one1=j(n1,1,1); one2=j(n2,1,1);
    ident1=i(n1); ident2=i(n2);
    ybar1=x1`*one1/n1; ybar2=x2`*one2/n2;
    s1=x1`*(ident1-one1*one1`/n1)*x1/(n1-1.0);
    s2=x2`*(ident2-one2*one2`/n2)*x2/(n2-1.0);
    spool=((n1-1.0)*s1+(n2-1.0)*s2)/(n1+n2-2.0);
    t2=(ybar1-ybar2)`*inv(spool*(1/n1+1/n2))*(ybar1-ybar2);
    f=(n1+n2-k-1)*t2/k/(n1+n2-2);
    df1=k; df2=n1+n2-k-1;
    p=1-probf(f,df1,df2);
    print t2 f df1 df2 p;
  finish;
  use milk;
    read all var{fuel repair capital} where (truck="gasoline") into x1;
    read all var{fuel repair capital} where (truck="diesel") into x2;
  run hotel2; quit;
%let p=3;
data gasoline;
  set milk;
  if truck="gasoline";
  variable="fuel"; x=fuel; output;
  variable="repair"; x=repair; output;
  variable="capital"; x=capital; output;
  keep truck variable x;  run;
proc sort;
  by variable;  run;
proc means noprint;
  by variable;  id truck;  var x;
  output out=pop1 n=n1 mean=xbar1 var=s21;
data fake;
  set milk;
  if truck="diesel";
  variable="fuel"; x=fuel; output;
  variable="repair"; x=repair; output;
  variable="capital"; x=capital; output;
  keep truck variable x;  run;
proc sort;
  by variable;  run;
proc means noprint;
  by variable;  id truck;  var x;
  output out=pop2 n=n2 mean=xbar2 var=s22;
data combine;
  merge pop1 pop2;
  by variable;
  f=finv(0.99,&p,n1+n2-&p-1);
  t=tinv(1-0.005/&p,n1+n2-2);
  sp=((n1-1)*s21+(n2-1)*s22)/(n1+n2-2);
  losim=xbar1-xbar2-sqrt(&p*(n1+n2-2)*f*(1/n1+1/n2)*sp/(n1+n2-&p-1));
  upsim=xbar1-xbar2+sqrt(&p*(n1+n2-2)*f*(1/n1+1/n2)*sp/(n1+n2-&p-1));
  lobon=xbar1-xbar2-t*sqrt((1/n1+1/n2)*sp);
  upbon=xbar1-xbar2+t*sqrt((1/n1+1/n2)*sp);  run;
proc print;
  var variable losim upsim lobon upbon;  run;
```

SAS code for Problem 2

```
data swiss;
  infile "v:\505\datasets\swiss.dat";
  input length left right bottom top diag;
  run;
proc corr data=swiss nocorr cov out=corr_out;
  run;
proc iml;
  use corr_out;
  read all var _NUM_ where(_TYPE_="MEAN") into xbar[colname=varnames];
  read all var _NUM_ where(_TYPE_="COV") into S;
  read all var _NUM_ where(_TYPE_="N") into n;
  varnames=t(varNames);
  s2=vecdiag(S);
  n=n[1];
  p=nrow(s);
  xbar=t(xbar);
  mu0={215,130,130,9,10,141.354};
  t2=n*t(xbar-mu0)*inv(s)*(xbar-mu0);
  f=t2*(n-p)/p/(n-1);
  pval=1-probf(f,p,n-p);
  print varnames xbar mu0;
  print t2 f pval;
  alpha = .05;
  tb=tinv(1-alpha/2/p,n-1);
  lobon=xbar-tb*sqrt(s2/n);
  upbon=xbar+tb*sqrt(s2/n);
  print varNames lobon mu0 upbon;
  quit;
%let p=6;
data swiss2;
  set swiss;
  variable="length"; ratio=length/215; output;
  variable="left"; ratio=left/130; output;
  variable="right"; ratio=right/130; output;
  variable="bottom"; ratio=bottom/9; output;
  variable="top"; ratio=top/10; output;
  variable="diag"; ratio=diag/141.354; output;
  keep variable ratio;
  run;
proc sort;
  by variable;  run;
proc means;
  by variable;
  var ratio;
  output out=hw6a n=n mean=xbar var=s2;
  run;
data hw6b;
  set hw6a;
  f=finv(0.95,&p,n-&p);
  ratio=xbar; output;
  ratio=xbar-sqrt(&p*(n-1)*f*s2/(n-&p)/n); output;
  ratio=xbar+sqrt(&p*(n-1)*f*s2/(n-&p)/n); output;
  run;
proc gplot data=hw6b;
  axis1 length=4in;
  axis2 length=4in;
  plot ratio*variable / vaxis=axis1 haxis=axis2 vref=1 lvref=20;
  symbol v=none i=hilot color=black;
  run;
```