Some Large Sample Chi-squared Tests[1]

STA442/2101 Fall 2014

[1] See last slide for copyright information.

Overview

Large-Sample Chi-square

Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then recall

$$(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(p)$$

It's true asymptotically too.

Using $(\mathbf{X}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2(p)$

Suppose

- $\sqrt{n}\left(\mathbf{T}_n - \boldsymbol{\theta}\right) \stackrel{d}{\rightarrow} \mathbf{T} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$ and
- $\widehat{\boldsymbol{\Sigma}}_n \stackrel{p}{\rightarrow} \boldsymbol{\Sigma}$.

Then approximately as $n \rightarrow \infty$, $\mathbf{T}_n \sim N\left(\boldsymbol{\theta}, \frac{1}{n}\boldsymbol{\Sigma}\right)$, and

$$
\begin{array}{rcl}
W_n & = & (\mathbf{T}_n - \boldsymbol{\theta})^\top \left(\frac{1}{n}\boldsymbol{\Sigma}\right)^{-1} (\mathbf{T}_n - \boldsymbol{\theta}) \sim \chi^2(p) \\
 & & \qquad\qquad || \\
 & & n\left(\mathbf{T}_n - \boldsymbol{\theta}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\mathbf{T}_n - \boldsymbol{\theta}\right) \\
 & \approx & n\left(\mathbf{T}_n - \boldsymbol{\theta}\right)^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \left(\mathbf{T}_n - \boldsymbol{\theta}\right) \\
 & \sim & \chi^2(p)
\end{array}
$$

Or we could be more precise

Suppose

- $\sqrt{n}\left(\mathbf{T}_n - \boldsymbol{\theta}\right) \stackrel{d}{\rightarrow} \mathbf{T} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$ and
- $\widehat{\boldsymbol{\Sigma}}_n \stackrel{p}{\rightarrow} \boldsymbol{\Sigma}$.

Then $\widehat{\boldsymbol{\Sigma}}_n^{-1} \stackrel{p}{\rightarrow} \boldsymbol{\Sigma}^{-1}$, and by a Slutsky lemma,

$$\left(\begin{array}{c} \sqrt{n}\left(\mathbf{T}_n - \boldsymbol{\theta}\right) \\ \widehat{\boldsymbol{\Sigma}}_n^{-1} \end{array} \right) \stackrel{d}{\rightarrow} \left(\begin{array}{c} \mathbf{T} \\ \boldsymbol{\Sigma}^{-1} \end{array} \right).$$

By continuity,

$$\begin{array}{rcl} W_n & = & \left(\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta})\right)^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \\ & = & n\left(\mathbf{T}_n - \boldsymbol{\theta}\right)^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \left(\mathbf{T}_n - \boldsymbol{\theta}\right) \\ & \stackrel{d}{\rightarrow} & \mathbf{T}^\top \boldsymbol{\Sigma}^{-1} \mathbf{T} \\ & \sim & \chi^2(p) \end{array}$$

If $H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{h}$ is true

Where $\mathbf{L}$ is $r \times p$ and of full row rank

Asymptotically, $\mathbf{LT}_n \sim N\left(\mathbf{L}\boldsymbol{\theta}, \frac{1}{n}\mathbf{L\Sigma L}^\top\right)$. So

$$
\begin{array}{rcl}
 & & (\mathbf{LT}_n - \mathbf{L}\boldsymbol{\theta})^\top \left(\frac{1}{n}\mathbf{L\Sigma L}^\top\right)^{-1} (\mathbf{LT}_n - \mathbf{L}\boldsymbol{\theta}) \sim \chi^2(r) \\
 & & \qquad\qquad || \\
 & & n\left(\mathbf{LT}_n - \mathbf{h}\right)^\top \left(\mathbf{L\Sigma L}^\top\right)^{-1} \left(\mathbf{LT}_n - \mathbf{h}\right) \\
 & \approx & n\left(\mathbf{LT}_n - \mathbf{h}\right)^\top \left(\mathbf{L}\widehat{\mathbf{\Sigma}}_n\mathbf{L}^\top\right)^{-1} \left(\mathbf{LT}_n - \mathbf{h}\right) \\
= W_n & \sim & \chi^2(r)
\end{array}
$$

Or we could be more precise and use Slutsky lemmas.

Test of $H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{h}$

Where $\mathbf{L}$ is $r \times p$ and of full row rank

$$W_n = n\left(\mathbf{L}\mathbf{T}_n - \mathbf{h}\right)^\top \left(\mathbf{L}\widehat{\boldsymbol{\Sigma}}_n\mathbf{L}^\top\right)^{-1}\left(\mathbf{L}\mathbf{T}_n - \mathbf{h}\right)$$

Distributed approximately as chi-squared with r degrees of freedom under H_0.

If $\mathbf{T}_n$ is the maximum likelihood estimator of $\boldsymbol{\theta}$, it's called a *Wald test* (and $\widehat{\boldsymbol{\Sigma}}_n$ has a special form).

Example: The `statclass` data

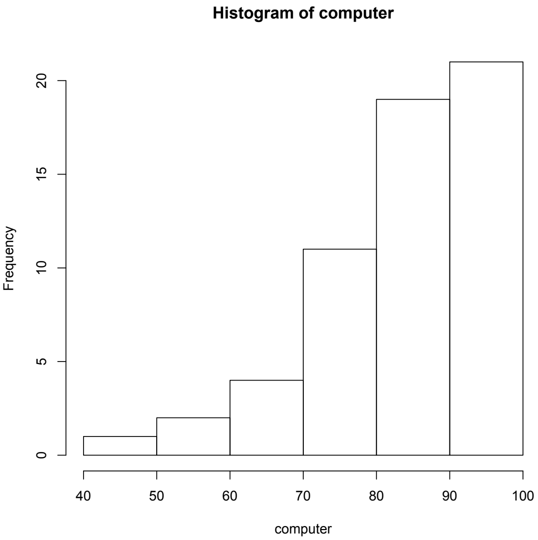
Fifty-eight students in a Statistics class took 8 quizzes, a midterm test and a final exam. They also had 9 computer assignments. The instructor wants to compare average performance on the four components of the grade.

- How about a model?
- Should we assume normality?
- Does it make sense to assume quiz marks independent of final exam marks?
- Does this remind you of a matched *t*-test?

Within cases versus between cases

- Want to compare average performance under several conditions, which are often experimental conditions, but not always.
- When a case (person, rat, school, etc.) appears in *all* the conditions, it's called a *within cases* design. Think of the matched t-test.
- When a case appears in *only one* condition, it's called a *between cases* design. Think of the two-sample t-test.
- Comparing performance on quizzes, midterm, final and computer assignments is within-cases.

Assume multivariate normality?



A model for the `statclass` data

Fifty-eight students in a Statistics class took 8 quizzes, a midterm test and a final exam. They also had 9 computer assignments.

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a random sample from an unknown distribution with mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)^\top$ and covariance matrix $\boldsymbol{\Sigma}$.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

Applying $W_n = n\left(\mathbf{LT}_n - \mathbf{h}\right)^\top \left(\mathbf{L}\widehat{\boldsymbol{\Sigma}}_n\mathbf{L}^\top\right)^{-1}\left(\mathbf{LT}_n - \mathbf{h}\right)$

To test $H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{h}$

- Test is based on $\sqrt{n}\left(\mathbf{T}_n - \boldsymbol{\theta}\right) \stackrel{d}{\rightarrow} \mathbf{T} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$
- CLT says $\sqrt{n}\left(\overline{\mathbf{Y}}_n - \boldsymbol{\mu}\right) \stackrel{d}{\rightarrow} \mathbf{Y} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$
- So $\mathbf{T}_n = \overline{\mathbf{Y}}_n$ and $\boldsymbol{\theta} = \boldsymbol{\mu}$.
- Sample variance-covariance matrix is good enough for $\widehat{\boldsymbol{\Sigma}}_n \stackrel{p}{\rightarrow} \boldsymbol{\Sigma}$
- Write $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ as $\mathbf{L}\boldsymbol{\mu} = \mathbf{h}$

$H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{h}$

To test equality of four means

$$
\begin{array}{cccc}
\mathbf{L} & \boldsymbol{\mu} & = & \mathbf{0} \\
\left(\begin{array}{rrrr} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{array}\right) & \left(\begin{array}{c} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{array}\right) & = & \left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right)
\end{array}
$$

Read the data

www.utstat.utoronto.ca/~brunner/appliedf14/code_n_data/lecture/statclass.data

S	E	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	C1	C2	C3	C4	C5	C6	C7	C8	C9	MT	Final
1	2	9	1	7	8	4	3	5	2	6	10	10	10	5	0	0	0	0	55	43
0	2	10	10	5	9	10	8	6	8	10	10	8	9	9	9	9	10	10	66	79
1	2	10	10	5	10	10	10	9	8	10	10	10	10	10	10	9	10	10	94	67
1	2	10	10	8	9	10	7	10	9	10	10	10	9	10	10	9	10	10	81	65
1	1	10	6	7	9	8	8	5	7	10	9	10	9	5	6	4	8	10	57	52
0	1	10	9	5	8	9	8	5	6	8	7	5	6	10	6	5	9	9	77	64
0	1	10	8	6	8	9	5	3	6	9	9	6	9	10	6	5	7	10	65	42
0	1	9	0	4	6	10	5	3	3	10	8	10	5	10	10	9	9	10	71	37
0	1	10	3	5	9	10	5	3	7	10	8	5	5	10	10	5	9	0	79	54
0	1	9	8	10	9	8	9	0	8	8	9	4	0	10	0	0	10	10	73	46
0	1	8	6	7	8	6	4	3	5	10	10	10	10	10	6	5	10	10	44	57
1	1	10	10	10	8	10	8	6	3	8	10	7	10	10	10	9	10	9	72	49

```
> statclass = read.table("http://www.utstat.toronto.edu/~brunner/appliedf14/cod
> head(statclass)
  S E Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 C1 C2 C3 C4 C5 C6 C7 C8 C9 MT  Final
1 1 2  9  1  7  8  4  3  5  2  6 10 10 10  5  0  0  0  0 55 43
2 0 2 10 10  5  9 10  8  6  8 10 10  8  9  9  9  9 10 10 66 79
3 1 2 10 10  5 10 10 10  9  8 10 10 10 10 10 10  9 10 10 94 67
4 1 2 10 10  8  9 10  7 10  9 10 10 10  9 10 10  9 10 10 81 65
5 1 1 10  6  7  9  8  8  5  7 10  9 10  9  5  6  4  8 10 57 52
6 0 1 10  9  5  8  9  8  5  6  8  7  5  6 10  6  5  9  9 77 64
> attach(statclass)
```

Process the data a bit

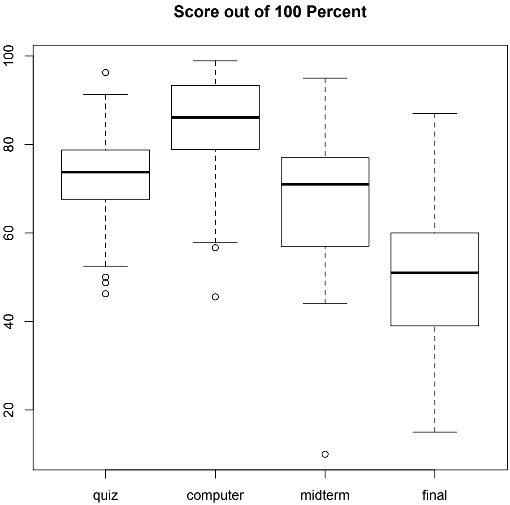
And take a look

```
> quiz = 10 * (Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8)/8
> computer = 10 * (C1+C2+C3+C4+C5+C6+C7+C8+C9)/9
> midterm = MT
> final = Final
> datta = cbind(quiz,computer,midterm,final); head(round(datta))

     quiz computer midterm final
[1,]   49       46      55    43
[2,]   82       93      66    79
[3,]   90       99      94    67
[4,]   91       98      81    65
[5,]   75       79      57    52
[6,]   75       72      77    64
```

```
> ybar = apply(datta,2,mean); ybar

    quiz computer  midterm    final
72.58621 83.98467 68.87931 49.44828
```

Boxplots

```
boxplot(datta); title("Score out of 100 Percent")
```

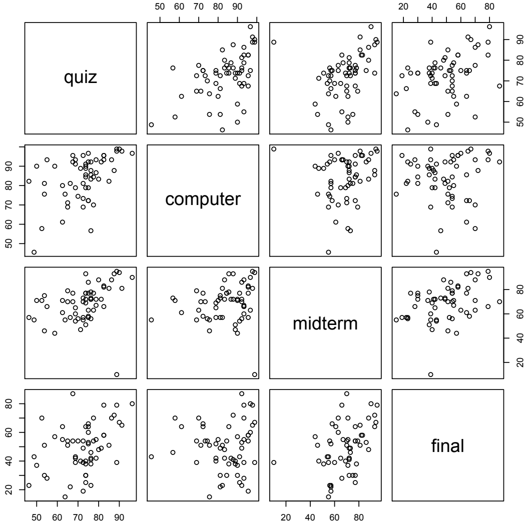


Covariances and Correlations

```
> sigmahat = var(datta); sigmahat

              quiz   computer   midterm      final
quiz     120.66130  62.369765  60.31760  71.736993
computer  62.36977 134.894281  27.68233   6.272098
midterm   60.31760  27.682328 223.37114  99.633999
final     71.73699   6.272098  99.63400 272.777979
```

```
> cor(datta)

              quiz   computer   midterm      final
quiz     1.0000000 0.48886995 0.3674063 0.39541626
computer 0.4888700 1.00000000 0.1594749 0.03269726
midterm  0.3674063 0.15947489 1.0000000 0.40363552
final    0.3954163 0.03269726 0.4036355 1.00000000
```

Scatterplot matrix



Calculate $W_n = n\left(\mathbf{L}\overline{\mathbf{T}}_n - \mathbf{h}\right)^\top \left(\mathbf{L}\widehat{\boldsymbol{\Sigma}}_n\mathbf{L}^\top\right)^{-1}\left(\mathbf{L}\overline{\mathbf{T}}_n - \mathbf{h}\right)$

To test $H_0 : \mathbf{L}\boldsymbol{\mu} = \mathbf{0}$

```
> L = rbind(c(1,-1,0,0),
+           c(0,1,-1,0),
+           c(0,0,1,-1) )
> n = length(quiz); n
[1] 58
> Wn = n * t(L %*% ybar) %*% solve(L%*%sigmahat%*%t(L)) %*% L%*%ybar
> Wn
         [,1]
[1,] 176.8238
> Wn = as.numeric(Wn)
> pvalue = 1-pchisq(Wn,df=3); pvalue
[1] 0
```

Conclude that the four means are not all equal. Which ones are different from one another? Need follow-up tests.

The R function Wtest

"Estimated" asymptotic covariance matrix $\widehat{\mathbf{V}}_n = \frac{1}{n}\widehat{\boldsymbol{\Sigma}}_n$

$$\begin{aligned} W_n &= n\,(\mathbf{L}\mathbf{T}_n - \mathbf{h})^\top \left(\mathbf{L}\widehat{\boldsymbol{\Sigma}}_n\mathbf{L}^\top\right)^{-1} (\mathbf{L}\mathbf{T}_n - \mathbf{h}) \\ &= (\mathbf{L}\mathbf{T}_n - \mathbf{h})^\top \left(\mathbf{L}\widehat{\mathbf{V}}_n\mathbf{L}^\top\right)^{-1} (\mathbf{L}\mathbf{T}_n - \mathbf{h}) \end{aligned}$$

```
Wtest = function(L,Tn,Vn,h=0) # H0: L theta = h
# Note Vn is the estimated asymptotic covariance matrix of Tn,
# so it's Sigma-hat divided by n. For Wald tests based on numerical
# MLEs, Tn = theta-hat, and Vn is the inverse of the Hessian.
     {
     Wtest = numeric(3)
     names(Wtest) = c("W","df","p-value")
     r = dim(L)[1]
     W = t(L%*%Tn-h) %*% solve(L%*%Vn%*%t(L)) %*%
          (L%*%Tn-h)
     W = as.numeric(W)
     pval = 1-pchisq(W,r)
     Wtest[1] = W; Wtest[2] = r; Wtest[3] = pval
     Wtest
     } # End function Wtest
```

Illustrate the `Wtest` function

For $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$, got $W_n = 176.8238, df = 3, p \approx 0$.

```
> V = sigmahat / length(final)
> # Asymptotic covariance matrix of Y-bar is Sigma/n
> LL = rbind( c(1,-1,0,0),
+             c(0,1,-1,0),
+             c(0,0,1,-1) )
> Wtest(LL,ybar,V)
       W       df  p-value
176.8238   3.0000   0.0000
>
> ybar
    quiz computer  midterm    final
72.58621 83.98467 68.87931 49.44828
```

Is average quiz score different from midterm?

```
> L1 = rbind(c(1,0,-1,0)); n = length(final)
> Wtest(L=L1,Tn=ybar,Vn=sigmahat/n)
         W          df     p-value
3.56755878 1.00000000 0.05891887
```

Test the other pairwise differences between means.

Another application: Mean index numbers

In a study of consumers' opinions of 5 popular TV programmes, 240 consumers who watch all the shows at least once a month completed a computerized interview. On one of the screens, they indicated how much they enjoyed each programme by mouse-clicking on a 10cm line. One end of the line was labelled "Like very much," and the other end was labelled "Dislike very much." So each respondent contributed 5 ratings, on a continuous scale from zero to ten.

The study was commissioned by the producers of one of the shows, which will be called "Programme E." Ratings of Programmes A through D were expressed as percentages of the rating for Programme E, and these were described as "Liking indexed to programme E."

In statistical language

We have $X_{i,1}, \ldots X_{i,5}$ for $i = 1, \ldots, n$, and we calculate

$$Y_{i,j} = 100 \frac{X_{i,j}}{X_{i,5}}$$

- We want confidence intervals for the 4 mean index numbers, and tests of differences between means.
- Observations from the same respondent are definitely not independent.
- What is the distribution?
- What is a reasonable model?

Model

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a random sample from an unknown multivariate distribution F with expected value $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

One way to think about it is

- The parameter is the unknown distribution F.
- The parameter space is a space of distribution functions.
- $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are *functions* of F.
- We're only interested in $\boldsymbol{\mu}$.

We have the tools we need

- $\sqrt{n}(\overline{\mathbf{Y}}_n - \boldsymbol{\mu}) \stackrel{d}{\rightarrow} \mathbf{Y} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and
- For $\widehat{\boldsymbol{\Sigma}}_n \stackrel{p}{\rightarrow} \boldsymbol{\Sigma}$, use the sample covariance matrix.
- $H_0 : \mathbf{L}\boldsymbol{\mu} = \mathbf{h}$

$$W_n = n\left(\mathbf{L}\overline{\mathbf{Y}}_n - \mathbf{h}\right)^\top \left(\mathbf{L}\widehat{\boldsymbol{\Sigma}}_n\mathbf{L}^\top\right)^{-1}\left(\mathbf{L}\overline{\mathbf{Y}}_n - \mathbf{h}\right)$$

Read the data

```
> Y = read.table("http://www.utstat.toronto.edu/~brunner/appliedf14/code_n_data/lecture/TVshows.data")

> Y[1:4,]
      A     B     C     D
1 101.3  81.0 101.8  89.6
2  94.0  85.3  76.3 100.8
3 145.4 138.7 151.0 148.3
4  72.0  86.1  96.1  96.3
> n = dim(Y)[1]; n
[1] 240
```

Confidence intervals: $\overline{Y} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}$

```
> ave = apply(Y,2,mean); ave
        A         B         C         D
101.65958  98.50167  99.39958 103.94167

> v = apply(Y,2,var) # Sample variances with n-1
> stderr = sqrt(v/n)
> me95 = 1.96*stderr
> lower95 = ave-me95
> upper95 = ave+me95
> Z = (ave-100)/stderr
> rbind(ave,marginerror95,lower95,upper95,Z)
                       A          B           C          D
ave           101.659583  98.501667  99.3995833 103.941667
marginerror95   1.585652   1.876299   1.7463047   1.469928
lower95       100.073931  96.625368  97.6532786 102.471739
upper95       103.245236 100.377966 101.1458880 105.411594
Z               2.051385  -1.565173  -0.6738897   5.255814
```

What if we "assume" normality and use t?

```
> rbind(ave,lower95,upper95,Z)
                 A          B           C          D
ave     101.659583  98.501667  99.3995833 103.941667
lower95 100.073931  96.625368  97.6532786 102.471739
upper95 103.245236 100.377966 101.1458880 105.411594
Z         2.051385  -1.565173  -0.6738897   5.255814
> attach(Y) # So A, B, C, D are available
> t.test(A,mu=100)

One Sample t-test

data:  A
t = 2.0514, df = 239, p-value = 0.04132
alternative hypothesis: true mean is not equal to 100
95 percent confidence interval:
 100.0659 103.2533
sample estimates:
mean of x
 101.6596
```

Test equality of means

```
> S = var(Y); S
         A         B         C        D
A 157.0779 110.77831 106.56220 109.6234
B 110.7783 219.93950  95.66686 100.3585
C 106.5622  95.66686 190.51937 106.2501
D 109.6234 100.35851 106.25006 134.9867
> cor(Y)
          A         B         C         D
A 1.0000000 0.5959991 0.6159934 0.7528355
B 0.5959991 1.0000000 0.4673480 0.5824479
C 0.6159934 0.4673480 1.0000000 0.6625431
D 0.7528355 0.5824479 0.6625431 1.0000000
>
> L4 = rbind( c(1,-1, 0, 0),
+             c(0, 1,-1, 0),
+             c(0, 0, 1,-1)  )
> Wtest(L=L4,Tn=ave,Vn=S/n)
           W           df      p-value
7.648689e+01 3.000000e+00 2.220446e-16
```

Pairwise comparisons
Where is the effect coming from?

Set it up.

```
> testmatrix = diag(1,4,4) # Start with an identity matrix.
> labelz = colnames(Y)
> rownames(testmatrix) = labelz; colnames(testmatrix) = labelz
> testmatrix
```

```
  A B C D
A 1 0 0 0
B 0 1 0 0
C 0 0 1 0
D 0 0 0 1
```

Fill the matrix

```
> for(i in 1:3)
+     {
+     for(j in (i+1):4)
+         {
+         LL = rbind(c(0,0,0,0))
+         LL[i]=1; LL[j]=-1
+         print(LL) # Just to check
+         W = Wtest(L=LL,Tn=ave,Vn=S/n)
+         testmatrix[i,j] = W[1]; testmatrix[j,i]=W[3]
+         } # Next j
+     } # Next i
     [,1] [,2] [,3] [,4]
[1,]    1   -1    0    0
     [,1] [,2] [,3] [,4]
[1,]    1    0   -1    0
     [,1] [,2] [,3] [,4]
[1,]    1    0    0   -1
     [,1] [,2] [,3] [,4]
[1,]    0    1   -1    0
     [,1] [,2] [,3] [,4]
[1,]    0    1    0   -1
```

Look at the $\binom{4}{2}$ pairwise comparisons

```
> # Test statistics (chisq with 1 df) are in the upper triangle,
> # p-values in lower
> round(testmatrix,4)

       A       B      C       D
A 1.0000 15.3954 9.1158 17.1647
B 0.0001  1.0000 0.8831 46.0573
C 0.0025  0.3474 1.0000 43.8147
D 0.0000  0.0000 0.0000  1.0000
```

```
> ave

        A         B         C         D
101.65958  98.50167  99.39958 103.94167
```

Average reported enjoyment was greatest for Program D, followed by A. The results are consistent with no difference between B and C.

Multiple Comparisons
The problem

- Most hypothesis tests are designed to be carried out in isolation.
- But if you do a lot of tests and all the null hypotheses are true, the chance of rejecting at least one of them can be a lot more than α. This is inflation of the Type I error probability.
- Otherwise known as the curse of a thousand t-tests.
- Multiple comparison procedures (sometimes called follow-up tests, post hoc tests, probing) try to offer a solution.

Multiple Comparisons
A solution

- Protect a *family* of tests against Type I error at some *joint* significance level α.
- If all the null hypotheses are true, the probability of rejecting at least one is no more than α.
- Many methods are available; we'll consider just one for now: Bonferroni.

Bonferroni multiple comparisons

- Based on Bonferroni's inequality:

$$Pr\left\{\cup_{j=1}^{k} A_j\right\} \leq \sum_{j=1}^{k} Pr\{A_j\}$$

- Applies to any collection of k tests.
- Assume that all k null hypotheses are true.
- Event A_j is that null hypothesis j is rejected.
- Do the tests as usual.
- Adjust the significance level, and reject each H_0 if $p < \alpha/k$.

$$Pr\left\{\cup_{j=1}^{k} A_j\right\} \leq \sum_{j=1}^{k} Pr\{A_j\} = \sum_{j=1}^{k} \alpha/k = \alpha$$

- Or, adjust the p-values. Multiply them by k, and reject if $pk < \alpha$.

TV show example

	A	B	C	D
A	1.0000	15.3954	9.1158	17.1647
B	0.0001	1.0000	0.8831	46.0573
C	0.0025	0.3474	1.0000	43.8147
D	0.0000	0.0000	0.0000	1.0000

- There are $\binom{4}{2} = 6 = k$ tests in the family.
- Adjusted α is $0.05/6 = 0.0083$.
- Conclusions don't change in this case.
- What if the family includes comparisons with Program E? Now there are 10 comparisons and H_0 is rejected if $p < \alpha/10 = 0.005$.

Include Z tests for comparison with Program E

Adjusted significance level is $\alpha/10 = 0.005$

```
> pval = 2*pnorm(-abs(Z))
> rbind(Z,pval)
```

```
              A          B          C            D
Z    2.05138485 -1.5651734 -0.6738897 5.255814e+00
pval 0.04022948  0.1175423  0.5003815 1.473709e-07
```

Add to the conclusions: Program D is preferred to E, but E is in a statistical tie with A, B and C.

Advantages and disadvantages

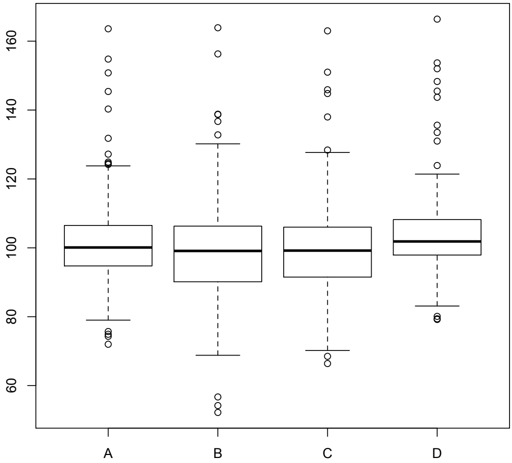
Of the Bonferroni method

- Advantage: Flexible Applies to any collection of hypothesis tests.
- Advantage: Easy to do.

- Disadvantage: Must know what all the tests are before seeing the data. So we were cheating.
- Disadvantage: A little conservative; the true joint significance level is less than α.

Practical versus statistical significance

`boxplot(Y)`



Between cases: Independent groups

Like a one-factor ANOVA

- Have n cases, separated into p groups: Maybe experimental treatment (say, drug) or occupation of main wage earner in family.
- $n_1 + n_2 + \cdots + n_p = n$
- Response variable is either binary or quantity of something, like annual energy consumption.
- No reason to believe normality.
- No reason to believe equal variances.
- $H_0 : \mathbf{L}\boldsymbol{\mu} = \mathbf{h}$
- For example, $H_0 : \mu_1 = \ldots = \mu_p$
- Or $\mu_2 = \mu_7$

Basic Idea

The p sample means are independent random variables. Asymptotically,

- $\overline{Y}_j \sim N(\mu_j, \frac{\sigma_j^2}{n_j})$
- The $p \times 1$ random vector $\overline{\mathbf{Y}}_n \sim N(\boldsymbol{\mu}, \mathbf{V}_n)$,
- Where $\mathbf{V}_n$ is a $p \times p$ diagonal matrix with jth diagonal element $\frac{\sigma_j^2}{n_j}$.
- $\mathbf{L}\overline{\mathbf{Y}}_n \sim N_r(\mathbf{L}\boldsymbol{\mu}, \mathbf{L}\mathbf{V}_n\mathbf{L}^\top)$
- "Estimate" $\mathbf{V}_n$ with the diagonal matrix $\widehat{\mathbf{V}}_n$, jth diagonal element $\frac{\widehat{\sigma}_j^2}{n_j}$
- And if $H_0 : \mathbf{L}\boldsymbol{\mu} = \mathbf{h}$ is true, then asymptotically

$$W_n = (\mathbf{L}\overline{\mathbf{Y}}_n - \mathbf{h})^\top \left(\mathbf{L}\widehat{\mathbf{V}}_n\mathbf{L}^\top\right)^{-1} (\mathbf{L}\overline{\mathbf{Y}}_n - \mathbf{h}) \sim \chi^2(r)$$

One little technical issue

- More than one n_j is going to infinity.
- The rates at which they go to infinity can't be too different.
- In particular, if $n = n_1 + n_2 + \cdots + n_p$,
- Then each $\frac{n_j}{n}$ must converge to a non-zero constant (in probability).

Loose asymptotic arguments lose this kind of detail.

Compare High School marks for students at 3 campuses

Campus	n	Mean	Standard Deviation
SG	3906	84.94	5.59
UTM	1583	79.68	5.82
UTSC	1849	79.96	5.98

Compute $W_n = \left(\mathbf{L}\overline{\mathbf{Y}}_n - \mathbf{h}\right)^\top \left(\mathbf{L}\widehat{\mathbf{V}}_n\mathbf{L}^\top\right)^{-1} \left(\mathbf{L}\overline{\mathbf{Y}}_n - \mathbf{h}\right)$

$H_0 : \mu_1 = \mu_2 = \mu_3$

Campus	n	Mean	Standard Deviation
SG	3906	84.94	5.59
UTM	1583	79.68	5.82
UTSC	1849	79.96	5.98

```
> source("http://www.utstat.utoronto.ca/~brunner/Rfunctions/Wtest.txt")

> n = c(3906,1583,1849)
> ybar = c(84.94,79.68,79.96)
> Vhat = diag(c(5.59,5.82,5.98)^2/n); Vhat
            [,1]      [,2]      [,3]
[1,] 0.008000026 0.0000000 0.0000000
[2,] 0.000000000 0.0213976 0.0000000
[3,] 0.000000000 0.0000000 0.0193404
> L1 = rbind(c(1,-1,0),
+            c(0,1,-1)  )
> Wtest(L1,ybar,Vhat)
      W      df  p-value
1441.58    2.00     0.00
```

Test difference between UTM and UTSC

Campus	n	Mean	Standard Deviation
SG	3906	84.94	5.59
UTM	1583	79.68	5.82
UTSC	1849	79.96	5.98

```
> # UTM vs. UTSC
> Wtest(rbind(c(0,1,-1)),ybar,Vhat)
        W        df   p-value
1.9244931 1.0000000 0.1653622
```

There are two more pairwise comparisons.

Copyright Information