MA231 – Spring 2009 **Homework 9** **Due 4/24/2009**

a.k.a. Homework Isham

Homework Policies: You should give a brief and concise explanation for each question. Just writing down an answer with no explanation is usually not sufficient. If the homework requires output from Stata, incorporate that output into your written assignments. Homework is due at the *beginning* of class on the day indicated.

(1) M&M 7.99, page 479

7.99. The test statistic is $F = \left(\frac{34.79}{33.24}\right)^2 \doteq 1.0954$, with df 70 and 36. The two-sided P-value is 0.7794, so we do not have enough evidence to conclude that the standard deviations are different. We do not know if the distributions are Normal, so this test may not be reliable. However, with s_1 and s_2 so close together, it seems likely that the conclusion ("we do not have enough evidence...") is appropriate; the reliability of the test would be a more crucial issue if we had rejected H_0.

(2) M&M 8.43, page 517

8.43. We have $\hat{p}_m = \frac{3547}{5594} \doteq 0.6341$, $\hat{p}_f = \frac{1447}{3469} \doteq 0.4171$, and pooled proportion $\hat{p} = \frac{3547+1447}{5594+3469} \doteq 0.5510$. For the test of H_0: $p_m = p_f$ vs. H_a: $p_m \neq p_f$, the appropriate standard error is $\text{SE}_{D_p} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{5594}+\frac{1}{3469}\right)} \doteq 0.01075$ and the test statistic is $z = (\hat{p}_m - \hat{p}_f)/\text{SE}_{D_p} \doteq 20.18$—overwhelming evidence that the two proportions are different. The standard error for a confidence interval is

$\text{SE}_D = \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2} \doteq 0.01056$, and the 95% confidence interval is 0.1962 to 0.2377. (The plus four interval is nearly identical: 0.1962 to 0.2376.) This interval and the significance test are only designed to deal with random sampling error; other sources of error such as nonresponse could throw all conclusions into doubt.

(3) M&M 8.77, page 523

8.77. The difference becomes more significant (i.e., the P-value decreases) as the sample size increases. For small sample sizes, the difference between $\hat{p}_1 = 0.5$ and $\hat{p}_2 = 0.4$ is not significant, but with larger sample sizes, we expect that the sample proportions should be better estimates of their respective population proportions, so $\hat{p}_1 - \hat{p}_2 = 0.1$ suggests that $p_1 \neq p_2$.

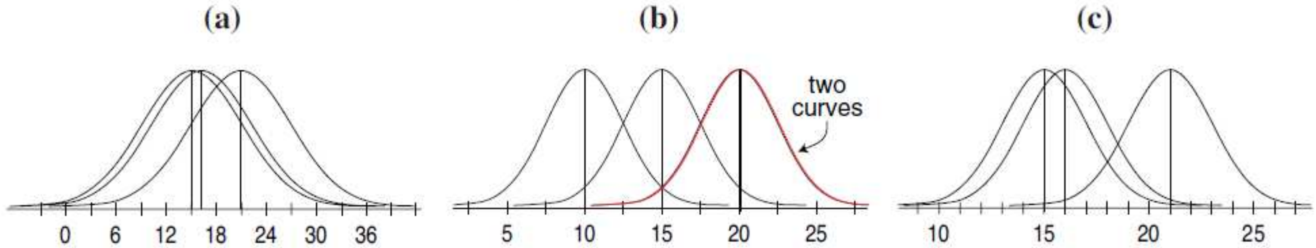
n	z	P
40	0.90	0.3681
50	1.01	0.3125
80	1.27	0.2041
100	1.42	0.1556
400	2.84	0.0045
500	3.18	0.0015
1000	4.49	0.0000

(4) M&M 12.2, page 644

12.2. **(a)** If we reject H_0, we conclude that *at least one* mean is different from the rest. **(b)** There is no theoretical limit to the number of means that can be compared. **(c)** Two-way ANOVA is used to examine the effect of two explanatory variables on a response variable.

(5) M&M 12.4 page 648

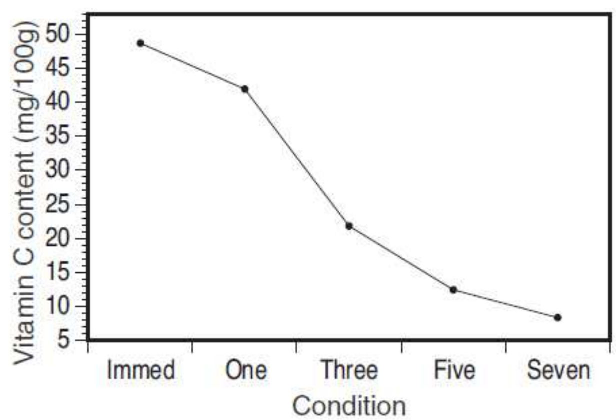
12.4. These pictures are shown below. Note that for (b), there are four curves, but two coincide almost exactly (one centered at 20, the other at 20.1).



(6) M&M 12.29, page 675

12.29. **(a)** Table below ($\bar{x}$, s, $s_{\bar{x}}$ in mg/100 g). **(b)** To test H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ vs. H_a: not all μ_i are equal, we have $F = 367.74$ with df 4 and 5, and $P < 0.0005$, so we reject the null hypothesis. Minitab output below. **(c)** Plot below. We conclude that vitamin C content decreases over time.

Condition	n	$\bar{x}$	s	$s_{\bar{x}}$
Immediate	2	48.705	1.534	1.085
One day	2	41.955	2.128	1.505
Three days	2	21.795	0.771	0.545
Five days	2	12.415	1.082	0.765
Seven days	2	8.320	0.269	0.190



Minitab output

```
Analysis of Variance on VitC
Source     DF        SS        MS        F        p
Days        4   2565.72    641.43   367.74    0.000
Error       5      8.72      1.74
Total       9   2574.44
```

(7) M&M 12.30, page 675

You only need to compare the immediate group to the 1-day group; and the immediate group to the 7-day group. Both are significant.

12.30. We have 10 comparisons to make, and df $= 5$, so the Bonferroni critical value with $\alpha = 0.05$ is $t^{**} = 4.7733$. The pooled standard deviation is $s_p \doteq 1.3207$, and the standard error of each difference is $s_p\sqrt{1/2 + 1/2} = s_p$, so two means are significantly different if they differ by $t^{**}s_p \doteq 6.3041$. All differences are significant *except* the five-day/seven-day difference.

STATA output

```
. oneway vitaminc condition, bon

                        Analysis of Variance
    Source              SS          df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      2565.72094      4   641.430234    367.74     0.0000
 Within groups      8.72120051      5    1.7442401
------------------------------------------------------------------------
    Total           2574.44214      9   286.049126

Bartlett's test for equal variances:  chi2(4) =   2.4331  Prob>chi2 = 0.657

              Comparison of Vitamin C Concentration by Condition
                                (Bonferroni)
Row Mean-|
Col Mean |   Immediat   One day    Three da   Five day
---------+--------------------------------------------
One day  |      -6.75
         |      0.037
         |
Three da |     -26.91     -20.16
         |      0.000      0.000
         |
Five day |     -36.29     -29.54      -9.38
         |      0.000      0.000      0.009
         |
Seven da |    -40.385    -33.635    -13.475     -4.095
         |      0.000      0.000      0.002      0.268
```

a.k.a. Homework Isham

Famous Statistician of the Week

Who is this?
Valerie Isham

Why is she cool?
Valerie Isham is a Professor in the Department of Statistical Science at University College London. Her research in applied probability involves the development and application of stochastic models. Her theoretical work contributes to the toolbox of tractable models available to an applied probabilist, and she is involved in interdisciplinary projects that apply such models to the physical and medical sciences.

"I use probability, or probability models, to represent the real world in some way. So I spend a lot of my time exploring models per se: thinking about assumptions that might be appropriate to a range of situations, and seeing what the consequences of those assumptions are."
Courtesy of *http://www.ma.hw.ac.uk/~ndg/fom/ishamqu.html*