Solutions to HW6, Stat 401 Spring 2011

(1). Ch. 14 #8. By my reading, category (1) consists of 15 days out of 365, (2) consists of 47 out of 365, (3) of 121 out of 365, and (4) of the remaining 182 out of 365. Thus the null-hypothesis $\mathbf{p}_0 = (15/365,\ 47/365,\ 121/365,\ 182/365)$, and the expected counts (obtained by multiplying $\mathbf{p}_0$ by 200, are $(8.22, 25.75, 66.30, 99.73)$ while the observed numbers are $(11, 24, 69, 96)$. The chi-square test statistic (3 degrees of freedom) is $2.78^2/8.22 + 1.75^2/25.75 + 2.7^2/66.3 + 3.73^2/99.73 = 1.31$, which has P-value 0.73, and we accept the null hypothesis at all reasonable α levels.

(2). Ch. 4 #92(a). Here we generate plots using the following `R` statements.

```
> Ex4.92 = scan("ex04-92.txt", skip=1)
> qqnorm(Ex4.92)
### normality very good; line with intercept  mean(Ex4.92)
#### and slope sd(Ex4.92) goes through almost all points.
```

(3). Ch. 10 #2. Here the data are as given in Ex.10.1, with $I = 4, J = 6$, except that, since all observations in Group 4 have 120 added to them, $\bar{X}_4 = 682.02$, $S_4 = 39.87$. Using these along with the previously given $\bar{X}_i$, S_i, we find: $SSTr = 6 * 3 * var(\{\bar{X}_i\}_{i=1}^4) = 18669.2$, $SSE = 4 * 5 * mean(\{S_i^2\}) = 33838.38$. Note that the $MSE = SSE/20 = 1691.92$ is the same as given on p. 375 in working with Ex. 10.1. The $F_{3,20}$ statistic for testing H_0 becomes $(18669.2/3)/(33838.4/20) = 3.678$ which has (one-sided) p-value `1 - pf(3.678,3,20)` $= .029$. So we reject H_0 at all α levels $> .029$.

(4). Ch. 10 #8. Here $I = 5, J = 7,$ and beginning with

```
> Ex10.8=array(read.table("ASCII/CH10/ex10-08.TXT",header=T,
sep=",")[,1],c(7,5), dimnames=list(1:7,c(4,6,8,10,12)))
> Xbar= apply(Ex10.8,2,mean); Ssq = apply(Ex10.8,2,var)
```

we calculate the entries of the ANOVA table as follows

Source	d.f.	SSQ	MSQ	F
Tr	4	43992.6	10998.2	10.48
Error	30	31475.0	1049.2	*
Total	34	75467.6	*	*

In doing this, it is a good idea to check that the given value of $\sum_{i,j} X_{ij}^2$ is equal to the value given. The F-statistic with 4 and 30 degrees of freedom gives 10.48, which has p-value 2.e-5. Thus we certainly reject at $\alpha = .01$.

(5). Ch. 10 #12. With $I = 5, J = 4$, we are given $MSE = 272.9$ and we calculate (using the new value $\bar{X}_{3\cdot}$.) $SSTr = 27929.0$, $MSTr = 6982.25$. We compare the differences between means with the quantity `qtukey(.95,5,15)`$*sqrt(272.9/4) = 3607$. So the group numbers which are found not significantly different (and would therefore be underscored together) are: (3,1), (1,4), (2,5). All other pairs of groups are found to be significantly different by the Tukey test. Although you were not required to calculate it here, the F statistic (with 4,15 df's) has the very large and significant value of 25.6.

(6). Ch. 10 #16. We are given results here in MINITAB output format.
(a). The standard deviations differ here by ratios as much as $44.51/20.83 = 2.14$. But based on $I = 5, J = 7$, we have sample sizes of $J = 7$ within each group, and an F-test for equality of variances in any pair of groups would compare the $F_{6,6}$ statistics (which are at most $2.14^2 = 4.57$) with $F_{6,6,.05} = 4.28$. The p-value for 4.57 is .043, which is **not** very extreme considering that it is the **most** extreme value out of $\binom{5}{2} = 10$ comparisons. Note that this too is a multpile-comparisons argument.
(b). The null hypothesis of no difference between means is highly significant becuase of the very small p-value for the $F_{4,30}$-statistic.
(c). The displayed 'critical value' for the Tukey test is 4.10 =`qtukey,.95,5,30)`. The displayed intervals are 95% CI's for pairwise differences, and those differences whose CI's do not contain 0 are significant. The only significant differences are those between groups with "level" identifiers 4 vs. 10, 4 vs. 12, 6 vs. 12, and 8 vs. 12. The mean-difference which is sufficiently large to be called 'significant' is $4.10 * sqrt(1049/7) = 50.19$.