Stat 371-003, Solutions to Homework #9

1. **10.50 (pg 433)**

(a) We refer to a χ^2 distribution with $(3-1)\times(3-1) = 4$ degrees of freedom. The critical value for $\alpha = 0.01$ (from Table 9) is 13.28, and so we *reject* the null hypothesis and conclude that the treatment influences claw configuration.

The P-value (from Table 9) is < 0.0001. With R, we find that the p-value is about 0.00002. (I used `1 - pchisq(24.36, 3)`.)

(b) The expected counts under the null hypothesis are as follows:

	Claw configuration			
Treatment	Right Crusher, Left Cutter	Right Cutter, Left Crusher	Right Cutter, Left Cutter	*Total*
Oyster chips	4.6	5.9	7.5	18
Smooth plastic	6.6	8.5	10.9	26
One oyster chip	5.8	7.6	9.6	23
Total	17	22	28	67

For example, the expected count for the first row, first column is $18 \times 17/67 \approx 4.6$.

The χ^2 statistic is then

$$\frac{(4.6-8)^2}{6.8} + \frac{(5.9-9)^2}{5.9} + \cdots + \frac{(9.6-7)^2}{7} \approx 24.36$$

(c) To get the percentage distribution for each of the three treatments, we divide the numbers in each row by the row sum.

	Claw configuration			
Treatment	Right Crusher, Left Cutter	Right Cutter, Left Crusher	Right Cutter, Left Cutter	*Total*
Oyster chips	44.4	50.0	5.6	100%
Smooth plastic	7.7	15.4	76.9	100%
One oyster chip	30.4	39.1	30.4	100%

(d) The smooth plastic treatment results in a high percentage of individuals both cutters. The oyster chips treatment results in a high percentage of individuals with a crusher.

2. **11.2 (pg 475)**

(a) The overall average is $(25*4+15*3+19*5)/(4+3+5) = 20$.

The between group sum of squares is

$$4\cdot(25-20)^2 + 3\cdot(15-20)^2 + 5\cdot(19-2)^2 = 180$$

The within group sum of squares is

$$(23-25)^2+(29-25)^2+(25-25)^2+(23-25)^2+(18-15)^2+\cdots+(19-19)^2 = 72$$

(b) The total sum of squares is

$$(23-20)^2+(29-20)^2+(25-20)^2+(23-20)^2+(18-20)^2+\cdots+(19-20)^2 = 252$$

And so we have 180 + 72 = 252.

(c) MS(between) = SS(between)/2 = 180/2 = 90.
MS(within) = SS(within)/9 = 72/9 = 8.
$s_{pooled} = \sqrt{\text{MS(between)}} = \sqrt{8} \approx 2.83$

3. **11.6 (pg 476)**

(a) The completed table is the following.

Source	**df**	**SS**	**MS**
Between groups	**3**	258	**86.0**
Within groups	26	**640**	**24.6**
Total	29	898	

(We fill in df and SS so that between + within = total; we calculate MS = SS/df just for the between and within rows.)

(b) The number of groups must be 3+1 = **4**.

(c) The total number of observations must be 29+1 = **30**.

4. **11.8 (pg 481)**

(a) The null hypothesis appears to be false. MAO activity appears higher for diagnosis I (chronic undifferentiated schizophrenic).

(b) The ANOVA table is as follows.

Source	**df**	**SS**	**MS**	**F**	**P-value**
Between groups	2	136.12	68.06	6.35	0.004
Within groups	39	418.25	10.72		
Total	41	554.37			

If we use Table 10 (looking at 2 and 40 df; 40 being close to 39) to get at the P-value, we'd find that the observed F statistic falls between 5.18 and 8.25, and so the P-value is between 0.001 and 0.01. Or we could use R by typing `1 - pf(6.35, 2, 39)`. With $\alpha = 0.05$, we reject the null hypothesis and conclude that there are differences among the three groups.

(c) $s_{\text{pooled}} = \sqrt{\text{MS(within)}} = \sqrt{10.72} \approx 3.27.$

5. Regarding the genenstein data, we read the data in as follows.

```
dat <- read.csv("http://www.biostat.wisc.edu/%7Ekbroman/teaching/stat371/genenstein.csv")
```

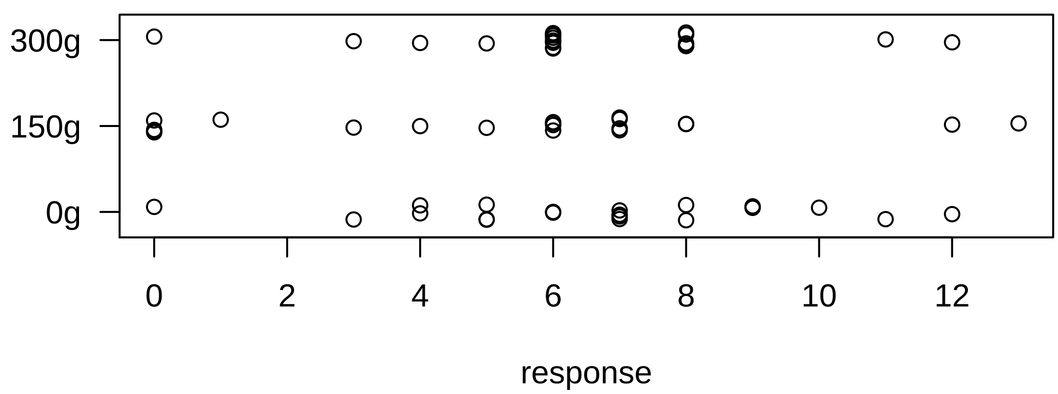
The column with the treatment groups is called `genenstein`. The column with the responses is called `response`.

(a) We can create the stripchart as follows.

```
stripchart(response ~ genenstein, data=dat, method="jitter", pch=1)
```

I use `method="jitter"` to jitter the points vertically. I use `pch=1` to get circles rather than squares (`pch` = "plot character").

Here's the plot.



(b) The anova table is calculated as follows.

```
summary(aov(response ~ genenstein, data=dat))
```

We obtain the following results.

```
            Df Sum Sq Mean Sq F value Pr(>F)
genenstein   2  21.06   10.53  1.1065 0.3376
Residuals   58 551.89    9.52
```

Note particularly that the degrees of freedom for `genenstein` (for the between-group differences) is **2**. If your results have **1**, then `genenstein` was not being treated as a *factor*, and you would need to type

```
dat$genenstein <- as.factor(dat$genenstein)
```

and re-run the ANOVA.

(c) Our conclusion: with P $\approx$ 34%, we fail to reject the null hypothesis (of a difference among the treatment groups). We the data provide insufficient evidence to conclude that genenstein is prevents cataract.