```
Populating the interactive namespace from numpy and matplotlib
```

Homework #6 ¶

10.40

	Average	Outstanding	Poor	Tot
Most desirable	20	21	4	45
Good	25	3	36	64
Adequate	8	14	2	24
Undesirable	7	10	6	23
total	60	48	48	156

Pearson's chi-square test

$$\chi^2 = \sum_{i,j} (n_{ij} - \hat{E}_{ij})^2 / \hat{E}_{ij}$$

where n_{ij} are the numbers in the table and $\hat{E}_{ij} = \frac{(row\ \ i\ \ total)(column\ \ j\ \ total)}{grand\ \ total}$

```
Yields 50.5500113225 on this data set
```

df $= (r-1)(c-1) = 3 * 2 = 6$

Reject H_0 if $\chi^2 > \chi^2_\alpha$

And here, if $\alpha = 0.05$ (or do I have that backwards?) χ^2_α equals

```
12.591587243743978
```

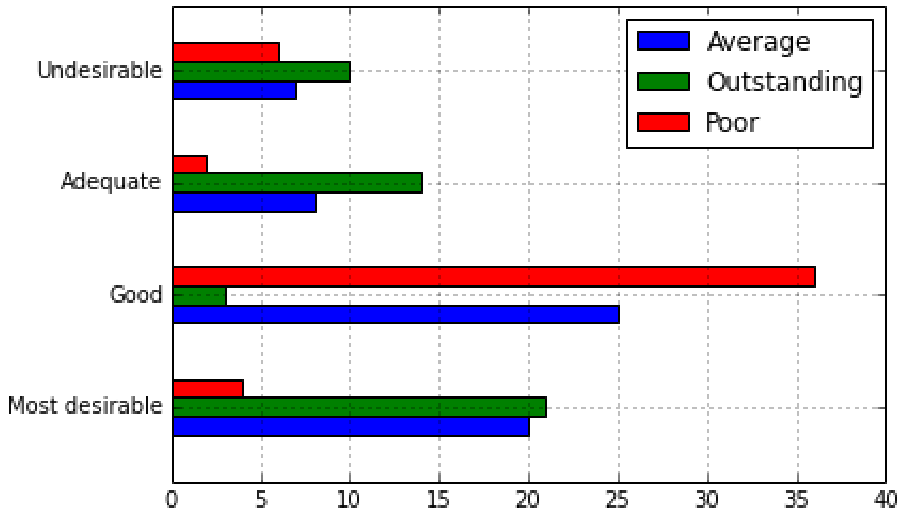
And therefore, because $\chi^2 = 50.55$ exceeds 12.59, H_0 is rejected. The $p - value = \mathbb{P}[\chi^2_6 > 50.55]$ equals

```
3.6466244468513764e-09
```

a.) the χ^2 equals 50.55 b.) the p-value is very small c.) Yes, because we rejected the null hypothesis d.) I have no idea

10.41

To me, simply visually looking at the data does not lead me to either accept or reject the research hypothesis



10.42

```
        Pearson's Chi-squared test

data:  df
X-squared = 13.0252, df = 3, p-value = 0.004582
```

	0	**1**	**2**	**3**
0	16	22.4	25.6	16
1	34	47.6	54.4	34

a.) Apparently the last row of ever cell is the expected number.

b.) $3 * 1 = 3$

c.) The p-value is small enough for us to reject the null hypothesis.

10.43

a.) Because the p-value is now over 90% we cannot reject the null-hypothesis.

b.) By combining the data, we now only have one degree of freedom. With only one degree, you must have some pretty big evidence for you to reject the null-hypothesis.

10.45

a.) By visual inspection of the data, you can see that there is evidence. Those three states have much higher adoption rates than the other three states... But at this point I feel so thoroughly bombarded with different test statistics, that I would not even know how to test this hypothesis.

b.)It provides justification because there is a higher rate of adoption amongst farmers who were provided information about the effectiveness of IPM.

8.6

a.) Visual inspection yields no evidence to indicate any difference

```
F =  4.849086119327985
```

```
	One-way analysis of means

data:  df2$a and df2$b
F = 4.847, num df = 3, denom df = 20, p-value = 0.01077
```

Using two different methods, I come to the conclusion that F is about 4.85.

```
b.) The critical value of F for alpha = 0.05 is  3.09839121214
Because the computed value of F exceeds this value, we reject the nu
ll hypothesis of equality of the mean.
However, since the two values are in the same order of magnitude, we
 are not surprised that a visual inspection of the data was inconclu
sive.
```

```
c.) p-value = 0.0107735195164
```

```
d.) Each of the poplations have a normal distribution, equal varianc
e and are randomly sampled.
e.) There could be a problem where one device randomly received too
many of one type of soil sample, thus throwing off the assumptions
```

8.9

$n_1 = n_2 = n_3 = 5$ are the sample sizes.

$t = 3$ is the number of populations.

apparently $\mu = 11$ and therefore $\tau_1 = -1.8$ $\tau_2 = -1.0$ $\tau_3 = 2.8$

which only leaves σ... but I'm not exactly sure how to calculate this. I will simple take the mean of the various variances: $\sigma = \sqrt{\frac{33.7+29+46.7}{3}} = 6.04$

8.12

residuals $= \epsilon_{ij} = y_{ij} - \bar{y}_i$

	1	2	3
count	5.000000	5.000000	5.00000
mean	9.600000	10.000000	13.80000
std	5.319774	5.385165	6.83374
min	4.000000	5.000000	8.00000
25%	5.000000	7.000000	9.00000
50%	10.000000	9.000000	12.00000
75%	12.000000	10.000000	15.00000
max	17.000000	19.000000	25.00000

Before we even begin, we notice that the book has a mistake in its calculations. The first sample mean, should be 9.6, not 9.2 as the book has written: $\frac{5+17+12+10+4}{5} = 9.6$

With that in mind, we calculate:

	0	1	2	3	4
1	-4.6	7.4	2.4	0.4	-5.6
2	9.0	0.0	-1.0	-3.0	-5.0
3	11.2	1.2	-1.8	-4.8	-5.8

b.) The sample variances are not equal, however they are not wildly different. So I would say that the AOV conditions are not atisfied, but one could probably run the analysis anyways.

8.30

```
	Kruskal-Wallis rank sum test

data:  df$a by df$b
Kruskal-Wallis chi-squared = 26.6213, df = 3, p-value = 7.068e-06
```

a.) This is a shift. The F value decreased by almost half a.) This is a shift. The P-value decreased by almost half.

b.) Even though there is a shift, the p-value remains close to zero, thus we reach the same conclusions as before.

8.32

```
        Kruskal-Wallis rank sum test

data:  df$a by df$b
Kruskal-Wallis chi-squared = 16.5614, df = 3, p-value = 0.0008698
```

```
        One-way analysis of means

data:  df$a and df$b
F = 11.0472, num df = 3, denom df = 28, p-value = 5.85e-05
```

a.) I did both an AOV test and a Kruskall-Wallis rank sum test.

b.) Both tests yield a p-value close to zero, thus we reject the null hypothesis and conclude that the different plots of land yield different amounts of corn.

8.33

errr... that's what I did. I still make the same conclusions.

8.36

No. The AOV conditions appear to be met.

The raw code for this IPython notebook is by default hidden for easier reading.

To toggle on/off the raw code, click here.