Supplementary exercise 9.50 of IPS7e

Data: Performance assessments (partial, usual, continuous; see text for further explanation) of employees in two age groups (under and above 40 years).

Model: Only little information is given about how the data were collected. It seems perhaps most natural that two samples of employees were taken, one for each age group. Then age group is not a response variable, it is instead an explanatory variable. Performance is clearly a response variable. Thus, the appropriate model would be two independent multinomial distributions, one for each column in the table (age group). This corresponds to assuming multinomial settings within each age group. It could also be argued that there probably is little interest in the age distribution among the employees, so that even if both variables were measured responses one may want to take the age groups as fixed. In that case, the analysis formally becomes conditional on age group.

Estimation: The proportions of main interest are within columns; the table below includes these proportions together with expected values and chi-square contributions for each cell.

```
MTB > XTabs 'perf' 'over40';
SUBC>   Layout 1 1;
SUBC>   Frequencies 'count';
SUBC>   Counts;
SUBC>   ColPercents;
SUBC>   ChiSquare;
SUBC>   Expected;
SUBC>   XResiduals;
SUBC>   DMissing 'perf' 'over40'.
```

Tabulated Statistics: perf, over40

Using frequencies in count

Rows: perf Columns: over40

	n	y	All
part	82	237	319
	16.33	31.14	25.26
	126.8	192.2	
	15.824	10.438	
usual	357	492	849
	71.12	64.65	67.22
	337.4	511.6	
	1.133	0.747	
contin	63	32	95
	12.55	4.20	7.52
	37.8	57.2	
	16.872	11.130	
All	502	761	1263
	100.00	100.00	100.00

Cell Contents
Count
% of Column
Expected count
Contribution to Chi-square

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	56.144	2	0.000
Likelihood Ratio	56.969	2	0.000

Test: The null hypothesis H_0 is that of equal performance distributions (i.e. same within-column proportions) in the two age groups, with a two-sided alternative H_a where there are (some) differences in the proportions between groups. The Pearson chi-square statistic (X^2) meets the conditions for its use because all expected values (in the listing) are well above 5. Minitab reports the results:

$$X^2 = 56.1, \quad \text{df} = (3-1)\cdot(2-1) = 2, \quad P\text{-value} < 0.0005.$$

The P-value is strongly significant and gives clear evidence of different performances in the groups of employees below and above 40 years of age.

Next we explore the parameter estimates, i.e. the sample proportions within each age group, to describe the differences. The values are quite close for the middle outcome ("usually") but differ considerably for the other outcome categories. The above-40 age group shows a higher proportion of performance assessments as "partial", and conversely the under-40 age group shows a higher proportion of "continually" performances. The contributions to the X^2 statistic confirm that these four cells are the ones of relevance to the significance achieved. In practical terms, the under-40 group received more assessments indicating continual exceedance of expectations than the above-40 group, and the over-40 group received more assessments indicating expectations to be (partially) met than the younger group. So the under-40 group was assessed more positively than the older folks.

Minitab technical note: In order to avoid reordering af rows and columns by their labels, one can set the value ordering of variables in the "Editor-Column properties-Value order" menu (available only when the worksheet is the active window, or by right-clicking a highlighted column of interest).