NHH, Bergen

Coding Errors - Solution

The data and its aggregate over the six branches are as follows:

Branch office name		A	B	C	D	E	F	All
Before training	No. records	258	220	293	159	462	389	1781
	No. errors	56	35	59	58	56	139	403
After training	No. records	409	332	420	219	444	882	2706
	No. errors	68	20	76	29	54	125	372

We want to compute and compare the fraction of errors before and after training, first for all records and then for each of the six branches separately. We get

$$\frac{403}{1781}-\frac{372}{2706}=0.226-0.137=0.089$$

We see that the error frequency in the total sample is reduced by 8.9% from 22.6% to 13.7%. This looks convincing in favour of the training program, i.e. is not likely to be due to chance. The computation as well as formal testing may be done by standard statistical software (here Minitab) as follows:

Test and CI for Two Proportions : All

```
Sample    X     N  Sample p
1       403  1781  0.226277
2       372  2706  0.137472

Difference = p (1) - p (2)
Estimate for difference:  0.0888051
95% CI for difference:  (0.0654395; 0.112171)
Test for difference = 0 (vs not = 0):  Z = 7.70  P-Value = 0.000
```

We see that the confidence interval (CI) with 95% guarantee of encompassing the true difference in error rates ranges from 6.5% to 11.2, i.e. the lower limit is far above zero. Alternatively we may look at the standard test for testing the hypothesis of zero difference between the error rates. The computed value of the test statistic is Z=7.70, which is far above the critical level for statistical significance, often taken as 2 which corresponds approximately to a 5% risk of false rejection of zero difference.

The formal analysis above assumes a common error probability for all records prior to training and all records after training. This is of course hardly justified. It may vary between the branches and vary between the employees within the branch. Nevertheless, the analysis above may be sufficient in practice to get an overall picture. However a lot more may be learnt by looking at the six branches separately. The results are as follows (computer output at the end)

Branch	A	B	C	D	E	F
Before %	21.7	15.9	20.1	36.4	12.1	35.7
After %	16.6	6.0	18.1	13.2	12.2	14.2
Difference %	5.1	9.9	2.0	23.2	-0.1	21.5
Z-statistic	1.64	3.80	0.68	5.13	-0.02	8.73
P-value	0.101	0.000	0.484	0.000	0.985	0.000

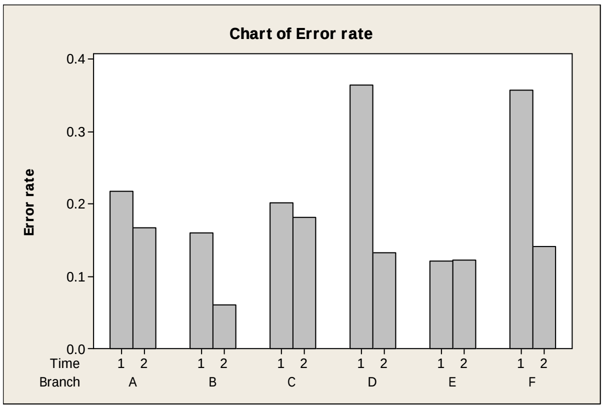
We see that the error rates before the training varies a lot, with E and B having far better results than A and C, which in turn are better than the D and F. The effect of the training varies as well. It obviously have had a substantial effect at the “inferior branches” D and F. For branch C and E there are no statistical significant effects. Since E is low at the outset this may lead one to believe that the training may not help getting further down. However, for branch B it has help to get down to 6% error, half the error rate of branch E. This reduction for B is highly statistical significant. For branch A there is a reduction as well, but it is not statistical significant, having a P-value of 10.1%. It could be argued that the test should be one-sided, based on the assumption that the training could never have a detriment effect. Then we may slice the P-value in half to 5.05%, which is close to being significant at the 5% significance level.

It is clearly of interest to get answers to some questions related to the revealed differences: Why was branch E better before the training? Could we learn from them? Why was the effect of the training at branch B so good? Why did the training not help branch C?

The investigation should be reviewed, so that we can rule out any differences in the set up, the training or the conditions for the data collection prior and after the training. It should also be worked out procedures so that what we have learned will have a lasting effect.

Note: More sophisticated modes of analysis are available, among then log-linear modelling. Here we see no good reason for this.

The following graph may be sufficient for communicating the results:



Computer output:

Test and CI for Two Proportions : A

```
Sample   X    N  Sample p
1       56  258  0.217054
2       68  409  0.166259


Difference = p (1) - p (2)
Estimate for difference:  0.0507951
95% CI for difference:  (-0.0111102; 0.112700)
Test for difference = 0 (vs not = 0):  Z = 1.64  P-Value = 0.101
```

Test and CI for Two Proportions : B

```
Sample   X    N  Sample p
1       35  220  0.159091
2       20  332  0.060241


Difference = p (1) - p (2)
Estimate for difference:  0.0988499
95% CI for difference:  (0.0441598; 0.153540)
Test for difference = 0 (vs not = 0):  Z = 3.80  P-Value = 0.000
```

Test and CI for Two Proportions : C

```
Sample   X    N  Sample p
1       59  293  0.201365
2       76  420  0.180952


Difference = p (1) - p (2)
Estimate for difference:  0.0204128
95% CI for difference:  (-0.0384430; 0.0792686)
Test for difference = 0 (vs not = 0):  Z = 0.68  P-Value = 0.494
```

Test and CI for Two Proportions : D

```
Sample   X    N  Sample p
1       58  159  0.364780
2       29  219  0.132420


Difference = p (1) - p (2)
Estimate for difference:  0.232360
95% CI for difference:  (0.145105; 0.319615)
Test for difference = 0 (vs not = 0):  Z = 5.30  P-Value = 0.000
```

Test and CI for Two Proportions : E

```
Sample   X    N  Sample p
1       56  462  0.121212
2       54  444  0.121622
```

```
Difference = p (1) - p (2)
Estimate for difference:  -0.000409500
95% CI for difference:  (-0.0429534; 0.0421344)
Test for difference = 0 (vs not = 0):  Z = -0.02  P-Value = 0.985
```

Test and CI for Two Proportions : F

```
Sample    X    N  Sample p
1       139  389  0.357326
2       125  882  0.141723


Difference = p (1) - p (2)
Estimate for difference:  0.215603
95% CI for difference:  (0.162711; 0.268495)
Test for difference = 0 (vs not = 0):  Z = 8.73  P-Value = 0.000

Fisher's exact test: P-Value = 0.00
```