NHH, Bergen

Tax Audit - Solution

(a)

By reading the report of the tax auditor you will see that the argument is mainly connected to a comparison of C with W and judging S/C by what is regarded an uncommon consumption, and when this occur. You realise that the role of the number of free guests (X) is largely neglected/underestimated, and that a more proper measure for the average consumption ought to be

$$R = \frac{S}{C + X}$$

The following four examples A-D represent realistic situations according to the text with respect to the number of cover charge tickets and the number of non-paying guests.

	A	B	C	D
(C, X)	(500, 1000)	(400, 200)	(600, 200)	(400, 400)
R	333	333	250	250
with $X = 0$	400	500	333	500
with $X = 300$	250	286	222	286

The calculations of the tax auditor (using $X = 0$) give a large mean consumption per guest in these situations, in two of them a extraordinary (and unreasonable?) high (B and D). The calculations where we take the number of possible non-paying guests into account show far less and more realistic average consumption. In the last example (D) the real consumption is in fact just half the one computed by the tax auditor.

The claim by the tax auditor is not justified, since the difference between high and low mean sales turnover depends critically of the value of $X,$ and its effect differ for different number of cover charge tickets, cf. situation B and C.

(b)

A possible defence is as follows:

«The average beer sale per cover charge ticket over the year 1997 was NOK 230, and over NOK 300 in only 5 evenings. The observed variation is in agreement with the operation of the establishment ”Nels”, with varying clientele and experience in this business, with large random variation of not easily identifiable causes. Furthermore, simple calculations (in (a)) show that if one accounts for a realistic number of non-paying guest, then the average beer sale will considerably less, and the amount will also show less variation . In this light the largest sales figures in comparison with the number of cover charge tickets is not strikingly large at all».

(c)
If we explain the number of cover charge ticket with the beer sale, which is undisputed, In a regression analysis, we see that the explanatory power is fairly good. Three evenings are exposed as outliers, with a small number of cover charge tickets compared to beer sales (negative St.Resid. marked by R). Only one of them is extreme (observation no. 100). This was an evening with an arrangement (cf. plot with black squares). It is a 5% risk for erroneously classifying an outlier. We found 5 (3 negative + 2 positive).This is in fact as expected among 104 observations!

An uncommonly large or small beer sale combined with uncommon C in comparison with this may tilt the regression line so that the line in total is misleading, and thereby hide some outliers. Five such instances are pointed out (2 small and 3 store, marked by X). From our computer output it may be difficult to tell whether this hides outliers.

(d)
Estimated percentages may be found from a regression analysis med January a basis (Coef Constant 0.95322 then means about 95%). The other months then come as an addition or subtraction from this. The result is given below in a table with rounded percentages. Among others we see that we get number grater than 100% . Correction for the number of non-paying guests may be done different ways. One possibility is to assume that the number of non-paying guests is a constant fraction of the number of paying guests. (How realistic this is may be discussed). A not unreasonable (and timid guess) is 1/3. We then get

$$\frac{W}{C+X}=\frac{W}{C+\frac{1}{3}C}=\frac{3}{4}\frac{W}{C},$$

i.e. the estimated % figures are scaled down by 75%. If we instead assume that the fraction is ½, we have to scale down the W/C by 67%. These two case are also presented in the table.

	J	F	M	A	M	J	J	A	S	O	N	D
%	95	101	80	72	62	49	42	37	53	102	93	95
75%	71	75	60	54	47	37	31	28	40	75	70	71
67%	63	67	53	48	41	32	28	24	35	68	62	63

We see that these ”corrected” figures are in well agreement with the claim by Nels Nelson, in particular the last line..

Computer output

Descriptive Statistics: Sales; Cover charge; S/C; W/C

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Sales	104	147252	148052	146815	30823	3022
Cover ch	104	654.2	646.0	652.9	160.8	15.8
S/C	104	230.34	221.50	227.89	40.32	3.95
W/C	104	0.7337	0.7339	0.7260	0.3023	0.0296

Variable	Minimum	Maximum	Q1	Q3
Sales	60747	224414	125496	166738
Cover ch	270.0	1112.0	524.5	794.0
S/C	161.00	386.00	202.00	253.75
W/C	0.1986	1.6325	0.4653	0.9416

Correlations: Sales; Cover charge; Wardrobe

	Sales	Cover charge
Cover charge	0.740	
Wardrobe	0.623	0.367

Regression Analysis: Cover charge versus Sales

The regression equation is
Cover charge = 86.1 + 0.00386 Sales

Predictor	Coef	SE Coef	T	P
Constant	86.05	52.27	1.65	0.103
Sales	0.0038580	0.0003475	11.10	0.000

S = 108.7   R-Sq = 54.7%   R-Sq(adj) = 54.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1456547	1456547	123.24	0.000
Residual Error	102	1205527	11819		
Total	103	2662074			

Unusual Observations

Obs	Sales	Cover ch	Fit	SE Fit	Residual	St Resid
8	165405	946.0	724.2	12.4	221.8	2.05R
25	76208	270.0	380.1	26.9	-110.1	-1.04 X
56	224414	1112.0	951.8	28.9	160.2	1.53 X
62	161290	1000.0	708.3	11.7	291.7	2.70R
80	221981	837.0	942.5	28.1	-105.5	-1.00 X
84	218040	801.0	927.3	26.8	-126.3	-1.20 X
89	132861	377.0	598.6	11.8	-221.6	-2.05R
95	163664	501.0	717.5	12.1	-216.5	-2.00R
100	173429	449.0	755.1	14.0	-306.1	-2.84R
103	60747	319.0	320.4	31.9	-1.4	-0.01 X

R denotes an observation with a large standardized residual
X denotes an observation whose X value gives it large influence.

Regression Analysis: W/C versus February; March; ...

The regression equation is
W/C = 0.953 + 0.0567 February - 0.153 March - 0.235 April - 0.338 May
- 0.460 June - 0.531 July - 0.585 August - 0.421 September
+ 0.0721 October - 0.0194 November + 0.0042 December

Predictor	Coef	SE Coef	T	P
Constant	0.95322	0.06737	14.15	0.000
February	0.05675	0.09821	0.58	0.565
March	-0.15273	0.09528	-1.60	0.112
April	-0.23504	0.09821	-2.39	0.019
May	-0.33780	0.09287	-3.64	0.000
June	-0.46040	0.09821	-4.69	0.000

July	-0.53080	0.09821	-5.40	0.000
August	-0.58547	0.09287	-6.30	0.000
Septembe	-0.42146	0.09821	-4.29	0.000
October	0.07210	0.09528	0.76	0.451
November	-0.01941	0.09528	-0.20	0.839
December	0.00416	0.09821	0.04	0.966

S = 0.2021 R-Sq = 60.1% R-Sq(adj) = 55.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	11	5.65706	0.51428	12.59	0.000
Residual Error	92	3.75840	0.04085		
Total	103	9.41546