Fox Module 16 analysis of variance: explanation of Duncan's prestige data

(The attached PDF file has better formatting.)

This file explains one-way analysis of variance on pages 147-148 of the Fox textbook. Final exam problems compute R^2 and *F*-statistics from the TSS, RSS, and RegSS.

Chapter 8 of the Fox textbook uses Duncan's prestige data to illustrate a one-way ANOVA analysis. The data shows 45 occupations, with four attributes:

- Type of occupation: professional or managerial, white collar, and blue collar.
- Average income
- Education
- Measure of prestige

Income, education, and prestige are scaled from 0 to 100.

The table below shows the data by occupation. The Excel workbook attached to this posting shows all the figures for the one-way analysis of variance. The computations are straight-forward; the final exam problems test these computations on a small data set.

Occupation	*Type*	*Income*	*Education*	*Prestige*
accountant	prof	62	86	82
pilot	prof	72	76	83
architect	prof	75	92	90
author	prof	55	90	76
chemist	prof	64	86	90
minister	prof	21	84	87
professor	prof	64	93	93
dentist	prof	80	100	90
reporter	wc	67	87	52
engineer	prof	72	86	88
undertaker	prof	42	74	57
lawyer	prof	76	98	89
physician	prof	76	97	97
welfare.worker	prof	41	84	59
teacher	prof	48	91	73
conductor	wc	76	34	38
contractor	prof	53	45	76
factory.owner	prof	60	56	81
store.manager	prof	42	44	45
banker	prof	78	82	92

bookkeeper	wc	29	72	39
mail.carrier	wc	48	55	34
insurance.agent	wc	55	71	41
store.clerk	wc	29	50	16
carpenter	bc	21	23	33
electrician	bc	47	39	53
RR.engineer	bc	81	28	67
machinist	bc	36	32	57
auto.repairman	bc	22	22	26
plumber	bc	44	25	29
gas.stn.attendantbc	bc	15	29	10
coal.miner	bc	7	7	15
streetcar.motorman	bc	42	26	19
taxi.driver	bc	9	19	10
truck.driver	bc	21	15	13
machine.operator	bc	21	20	24
barber	bc	16	26	20
bartender	bc	16	28	7
shoe.shiner	bc	9	17	3
cook	bc	14	22	16
soda.clerk	bc	12	30	6
watchman	bc	17	25	11
janitor	bc	7	20	8
policeman	bc	34	47	41
waiter	bc	8	32	10

The one-way ANOVA analysis tests whether the type of occupation affects prestige. Professional occupations have higher prestige than white collar or blue collar; we test if the differences are statistically significant.

Jacob: Prestige depends on education and income, not type of occupation. The highest blue collar prestige level (67) is for RR engineer, which also has the highest blue collar income (81). The highest professional prestige level is for physicians (97), who have the third highest education. In Duncan's study, dentists and lawyers have higher education, but this is probably measurement error: medical school along with internship and residency is longer for doctors than for dentists or lawyers.

Rachel: You are correct; the full ANOVA analysis considers also income and education. One-way ANOVA considers a simpler question: does prestige differ by type of occupation? Our goal is to explain the statistical technique. Education and income affect prestige, and this simple analysis is not complete.

The explanatory variable is type of occupation; the response variable is prestige. Regression analysis assumes the response variable has a normal distribution. But prestige is a value from 0 to 100; it does not

have a normal distribution. We transform prestige to *logit*(prestige / 100). The transformed response variable is closer to a normal distribution.

Jacob: How do we test if the response variable has a normal distribution?

Rachel: We use QQ plots. The QQ plot for prestige is thin-tailed; the QQ plot for *logit*(prestige/100) fits better to a normal distribution.

The textbook shows the computations for both prestige and logit(prestige/100).

The overall mean prestige is 47.68889. The mean prestige by type of occupation is

- Professional: 80.44444
- White collar: 36.66667
- Blue collar: 22.76190

The total sum of squares (TSS) is the square of the (prestige minus the average prestige). For accountants, this is $(82 - 47.68889)^2 = 1{,}177.25$

The residual sum of squares (RSS) is the square of the (prestige minus the average prestige for that type of occupation). For accountants, this is $(82 - 80.44444)^2 = 2.42$.

The regression sum of squares (RegSS) is the square of the (average prestige for the occupation minus the overall average prestige). For accountants, this is $(80.44444 - 47.68889)^2 = 1{,}072.93$.

Occupation	*Type*	*Inc*	*Edu*	*TSS*	*Prestige*	*Mn(prs)*	*RSS*	*RegSS*
accountant	prof	62	6	1,177.25	82	80.4444	2.42	1,072.92
pilot	prof	72	76	1,246.87	83	80.4444	6.53	1,072.93
architect	prof	75	92	1,790.23	90	80.4444	91.31	1,072.93
author	prof	55	90	801.52	76	80.4444	19.75	1,072.93
chemist	prof	64	86	1,790.23	90	80.4444	91.31	1,072.93
minister	prof	21	84	1,545.36	87	80.4444	42.98	1,072.93
professor	prof	64	93	2,053.10	93	80.4444	157.64	1,072.93
dentist	prof	80	100	1,790.23	90	80.4444	91.31	1,072.93
reporter	wc	67	87	18.59	52	36.6667	235.11	121.49
engineer	prof	72	86	1,624.99	88	80.4444	57.09	1,072.93
undertaker	prof	42	74	86.70	57	80.4444	549.64	1,072.93
lawyer	prof	76	98	1,706.61	89	80.4444	73.20	1,072.93
physician	prof	76	97	2,431.59	97	80.4444	274.09	1,072.93
welfare.worker	prof	41	84	127.94	59	80.4444	459.86	1,072.93
teacher	prof	48	91	640.65	73	80.4444	55.42	1,072.93
conductor	wc	76	34	93.87	38	36.6667	1.78	121.49
contractor	prof	53	45	801.52	76	80.4444	19.75	1,072.93
factory.owner	prof	60	56	1,109.63	81	80.4444	0.31	1,072.93
store.manager	prof	42	44	7.23	45	80.4444	1,256.31	1,072.93

banker	prof	78	82	1,963.47	92	80.4444	133.53	1,072.93
bookkeeper	wc	29	72	75.50	39	36.6667	5.44	121.49
mail.carrier	wc	48	55	187.39	34	36.6667	7.11	121.49
insurance.agen	wc	55	71	44.74	41	36.6667	18.78	121.49
store.clerk	wc	29	50	1,004.19	16	36.6667	427.11	121.49
carpenter	bc	21	23	215.76	33	22.7619	104.82	621.35
electrician	bc	47	39	28.21	53	22.7619	914.34	621.35
RR.engineer	bc	81	28	372.92	67	22.7619	1,957.01	621.35
machinist	bc	36	32	86.70	57	22.7619	1,172.25	621.35
auto.repairman	bc	22	22	470.41	26	22.7619	10.49	621.35
plumber	bc	44	25	349.27	29	22.7619	38.91	621.35
gas.stn.attenda	bc	15	29	1,420.45	10	22.7619	162.87	621.35
coal.miner	bc	7	7	1,068.56	15	22.7619	60.25	621.35
streetcar.motor	bc	42	26	823.05	19	22.7619	14.15	621.35
taxi.driver	bc	9	19	1,420.45	10	22.7619	162.87	621.35
truck.driver	bc	21	15	1,203.32	13	22.7619	95.29	621.35
machine.operat	bc	21	20	561.16	24	22.7619	1.53	621.35
barber	bc	16	26	766.67	20	22.7619	7.63	621.35
bartender	bc	16	28	1,655.59	7	22.7619	248.44	621.35
shoe.shiner	bc	9	17	1,997.10	3	22.7619	390.53	621.35
cook	bc	14	22	1,004.19	16	22.7619	45.72	621.35
soda.clerk	bc	12	30	1,737.96	6	22.7619	280.96	621.35
watchman	bc	17	25	1,346.07	11	22.7619	138.34	621.35
janitor	bc	7	20	1,575.21	8	22.7619	217.91	621.35
policeman	bc	34	47	44.74	41	22.7619	332.63	621.35
waiter	bc	8	32	1,420.45	10	22.7619	162.87	621.35
Total / average				43,687.64	47.6889		10,597.59	33,090.05

The total / average row shows that TSS (43,687.64) = RSS (10,597.59) + RegSS (33,090.05).

The prestige scores do not have a normal distribution. For a better ANOVA analysis, Fox uses the logit of the prestige scores divided by 100. Let Pr = prestige / 100, so *logit* (Pr) = *ln*(Pr) / (1 – *ln*(Pr)). We dot show the analysis of variance table for unadjusted prestige levels, though you can compute them easily from the last row of the table.

Logit of (Prestige / 100)

We form the same table using *logit* (prestige / 100). The attached Excel workbook has the same figures.

Occupation	*Type*	*I*	*E*	*TSS*	*Pres*	*logit(Pr)*	*Mn(pr)*	*RegSS*	*RSS*
accountant	prof	6	6	2.66960	82	1.5163	1.632114	3.06130	0.01340
pilot	prof	7	7	2.90079	83	1.5856	1.632114	3.06130	0.00216
architect	prof	7	9	5.35815	90	2.1972	1.632114	3.06130	0.31935
author	prof	5	9	1.61347	76	1.1527	1.632114	3.06130	0.22986
chemist	prof	6	8	5.35815	90	2.1972	1.632114	3.06130	0.31935
minister	prof	2	8	4.07435	87	1.9010	1.632114	3.06130	0.07228
professor	prof	6	9	7.31287	93	2.5867	1.632114	3.06130	0.91121
dentist	prof	8	1	5.35815	90	2.1972	1.632114	3.06130	0.31935
reporter	wc	6	8	0.03904	52	0.0800	-0.590384	0.22358	0.44947
engineer	prof	7	8	4.45199	88	1.9924	1.632114	3.06130	0.12983
undertaker	prof	4	7	0.15952	57	0.2819	1.632114	3.06130	1.82321
lawyer	prof	7	9	4.87652	89	2.0907	1.632114	3.06130	0.21034
physician	prof	7	9	12.91426	97	3.4761	1.632114	3.06130	3.40028
welfare.work	prof	4	8	0.23185	59	0.3640	1.632114	3.06130	1.60820
teacher	prof	4	9	1.23691	73	0.9946	1.632114	3.06130	0.40640
conductor	wc	7	3	0.13839	38	-0.4895	-0.590384	0.22358	0.01017
contractor	prof	5	4	1.61347	76	1.1527	1.632114	3.06130	0.22986
factory.owner	prof	6	5	2.45722	81	1.4500	1.632114	3.06130	0.03316
store.manag	prof	4	4	0.00691	45	-0.2007	1.632114	3.06130	3.35910
banker	prof	7	8	6.55304	92	2.4423	1.632114	3.06130	0.65648
bookkeeper	wc	2	7	0.10875	39	-0.4473	-0.590384	0.22358	0.02047
mail.carrier	wc	4	5	0.29784	34	-0.6633	-0.590384	0.22358	0.00532
insurance.ag	wc	5	7	0.06072	41	-0.3640	-0.590384	0.22358	0.05127
store.clerk	wc	2	5	2.37371	16	-1.6582	-0.590384	0.22358	1.14029
carpenter	bc	2	2	0.34886	33	-0.7082	-1.482151	1.86216	0.59902
electrician	bc	4	3	0.05650	53	0.1201	-1.482151	1.86216	2.56735
RR.engineer	bc	8	2	0.68183	67	0.7082	-1.482151	1.86216	4.79757
machinist	bc	3	3	0.15952	57	0.2819	-1.482151	1.86216	3.11171
auto.repairm	bc	2	2	0.86197	26	-1.0460	-1.482151	1.86216	0.19026
plumber	bc	4	2	0.60504	29	-0.8954	-1.482151	1.86216	0.34430
gas.stn.atten	bc	1	2	4.32508	10	-2.1972	-1.482151	1.86216	0.51133
coal.miner	bc	7	7	2.61488	15	-1.7346	-1.482151	1.86216	0.06373
streetcar.mot	bc	4	2	1.77547	19	-1.4500	-1.482151	1.86216	0.00103
taxi.driver	bc	9	1	4.32508	10	-2.1972	-1.482151	1.86216	0.51133

truck.driver	bc	2	1	3.18057	13	-1.9010	-1.482151	1.86216	0.17540
machine.ope	bc	2	2	1.07151	24	-1.1527	-1.482151	1.86216	0.10855
barber	bc	1	2	1.60973	20	-1.3863	-1.482151	1.86216	0.00919
bartender	bc	1	2	6.09668	7	-2.5867	-1.482151	1.86216	1.22000
shoe.shiner	bc	9	1	11.27990	3	-3.4761	-1.482151	1.86216	3.97583
cook	bc	1	2	2.37371	16	-1.6582	-1.482151	1.86216	0.03100
soda.clerk	bc	1	3	6.93792	6	-2.7515	-1.482151	1.86216	1.61134
watchman	bc	1	2	3.89351	11	-2.0907	-1.482151	1.86216	0.37038
janitor	bc	7	2	5.40471	8	-2.4423	-1.482151	1.86216	0.92198
policeman	bc	3	4	0.06072	41	-0.3640	-1.482151	1.86216	1.25034
waiter	bc	8	3	4.32508	10	-2.1972	-1.482151	1.86216	0.51133
Total / avg				134.1539	47.6	-0.1175		95.55014	38.60376

These tables show how Fox calculated the figures on page 148. Fox doesn't show all the work; the tables here show all the computations.

- The logit of Prestige / 100 for accountants is *ln*(0.82 / (1 – 0.82)) = 1.51635.
- The average logit for all occupations is –0.11754.
- The average logit by type of occupation is 1.6321 for professional, –0.5791 for white collar, and –14821 for blue collar.
- The regression sum of squares (RegSS) for accountants is $(1.6321 - -0.11754)^2 = 3.06124$.
- The residual sum of squares (RSS) for accountants is $(1.6321 - 1.5163)^2 = 0.01341$.

The total row in the table forms the analysis of variance. Fox shows the following table:

Source	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F*	*p*
Groups	95.550	2	47.775	51.98	<< 0.001
Residuals	38.604	42	0.919		
Total	134.154	44			

- The mean square is the sum of squares divided by the degrees of freedom.
  - 95.550 / 2 = 47.775; 38.604 / 42 = 0.9191.
- The F-statistic is the mean square for the groups divided by the mean square of the residuals.
  - 47.775 / 0.9191 = 51.98
- The R^2 is the sum of squares for the groups (RegSS) divided by the total sum of squares (TSS).
  - 95.550 / 134.154 = 71.22%

Jacob: How do we get the degrees of freedom?

Rachel: 45 data points minus 1 parameter (the mean) = 44 degrees of freedom for the total sum of squares.

Three occupation types (groups) minus one relation = 2 degrees of freedom for the groups. The relation is that an occupation is either professional, white collar, or blue collar.

Degrees of freedom for TSS – degrees of freedom for RegSS = degrees of freedom for RSS.