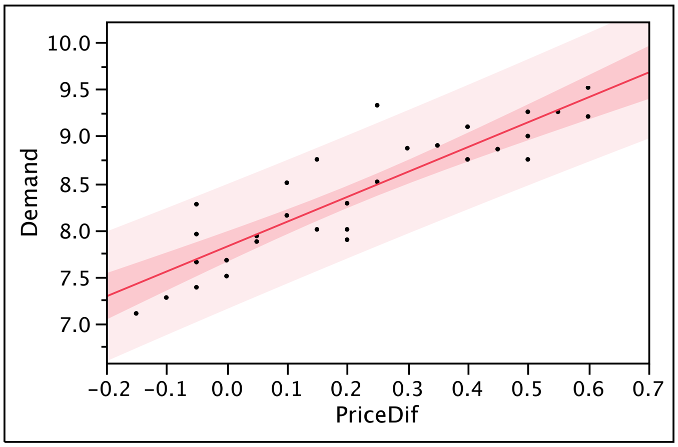
Insurance 260
Spring, 2009

Solutions, Assignment 1

3.9 This graphic shows the JMP output from the regression of sales (demand) in 100,000s of bottles of Fresh versus the price difference



Linear Fit

Demand = 7.8140876 + 2.6652145*PriceDif

Summary of Fit

RSquare	0.791516
RSquare Adj	0.78407
Root Mean Square Error	0.316561
Mean of Response	8.382667
Observations (or Sum Wgts)	30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	10.652685	10.6527	106.3028
Error	28	2.805902	0.1002	**Prob > F**
C. Total	29	13.458587		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	7.8140876	0.079884	97.82	<.0001*
PriceDif	2.6652145	0.2585	10.31	<.0001*

when compared to the industry average.

The line seems a reasonable summary of the relationship of demand to price difference. The growth seems linear, with comparable variance.

3.10

(a) The conditional mean $\mu_{y|0.10}$ is the mean demand in 100,000s of bottles when the price difference x_4 = 0.10, when Fresh costs 10 cents less than the industry average.

(b) The conditional mean $\mu_{y|-0.05}$ is the conditional mean demand in 100,000s of bottles when the price difference x_4 = -0.05, when Fresh costs 5 cents more than the industry average.

(c) The slope measures the average increase in demand per $1 difference in price. That's a bit of an extrapolation since most of the time the price difference is rather smaller.

(d) The intercept is the expected demand when Fresh sells the same as in the industry. That's well within the range of the data and a practical, easily interpreted value.

(e) Other things that influence demand, such as advertising, shelf placement in stores, numbers of stores where it is offered, packaging, and so forth.

3.11

(a) The estimated intercept is 7.81, implying that average demand (sales) is about 781,000 when the price difference is zero. The slope implies that on average, sales increase by about 27,000 per difference in price of \$0.10 (270,000 per dollar is a reach).

(b) These are the same, given by plugging in 0.10 into the regression equation,

$$\hat{y} = 7.814 + 2.665 * 0.10 = 808{,}050 \text{ bottles}$$

(c) This is not the best use of the equation. Nonetheless, by solving for the value of x that gives predicted value 850,000 we obtain (watch the decimal point)

$$8.5 = 7.814 + 2.665\,x \quad \text{implies} \qquad x = (8.5 - 7.814)/2.665 \approx 0.257$$

(d) You can see the quoted value for SSE on the output in the Analysis of Variance summary. The output also gives $s^2 = 0.1002$ and RMSE = s = 0.317 (about ± 31,700 bottles).

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	7.8140876	0.079884	97.82	<.0001*
PriceDif	2.6652145	0.2585	10.31	<.0001*

3.18

(a) Done in the previous exercise

(b) Done in the previous exercise

(c) This table shows the estimates and standard errors. The t-statistic (t-ratio) is computed by dividing the estimate by its standard error, as in 2.665/0.2585≈10.31

Term	Estimate	Std Error	Lower 95%	Upper 95%
Intercept	7.8140876	0.079884	7.650452	7.9777232
PriceDif	2.6652145	0.2585	2.1357021	3.1947269

(d) The p-value is far less than α=0.05. Reject H_0: $\beta_1 = 0$ and conclude that there is statistically significant linear association between price difference and demand (sales).

(e) Same as in "d". The p-value is smaller than any reasonable choice for α.

(f) Same as in "d".

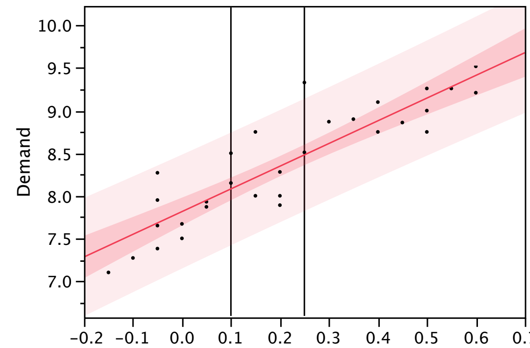
(g) The 95% confidence interval is approximately the estimate ± 2 standard errors. More precisely, use the 0.025 critical value from the t-distribution with 30-2 = 28 degrees of freedom, given by $t_{0.025,28} = 2.05$.

(h) JMP makes this tedious, so just use the t-distribution. The t-percentile is $t_{0.005,28} = 2.76$ so the interval is 2.665 ± 2.76(0.2585).

(i) See the table shown with "c" above. The t-ratio (t-stat) for the intercept is t = 7.81/0.07988 ≈ 97.82.

(j) As in testing the slope, the intercept is statistically significant at any reasonable α level.

(k) The standard error of the slope is
RMSE/sqrt(SSx) = 0.3166/sqrt((n-1)*var(x)) =



Demand	PriceDif	StdErr Pred Demand	Lower 95% Mean Demand	Upper 95% Mean Demand	StdErr Indiv Demand	Lower 95% Indiv Demand	Upper 95% Indiv Demand
•	0.1000000000	0.0648	7.9479	8.2133	0.3231	7.4187	8.7425
•	0.2500000000	0.0586	8.3604	8.6004	0.3219	7.8209	9.1398

0.3166/sqrt(29*0.0517)≈ 0.259
as shown in the JMP output.

3.22 I added 2 rows to the data table, then fit the simple regression using the "Fit Model" platform in order to be able to get the details of the prediction interval and confidence interval. You can also "eyeball" these from the plot of the fitted model. These are the upper and lower edges of the slices at x=0.10 and x=0.25 shown in the figure to the right.

(a) The 95% confidence interval at x = 0.10 is about 795,000 to 821,000. This interval is about the predicted value (Exercise 3.11) ± 2 times the SE of the predicted demand.

(b) The 95% prediction interval has the same center, but is wider to accommodate the additional period to period variation. Its about 743,000 to 874,000.

(c) See the picture.

(d) Use the t-percentile used to find the previous intervals. At x = 0.10 the predicted value is about 808,000. The interval for the mean is then

808,000 ± 2.76 (0.0648)

The 99% prediction interval for the individual value is then much wider at

808,000 ± 2.76 (0.323)

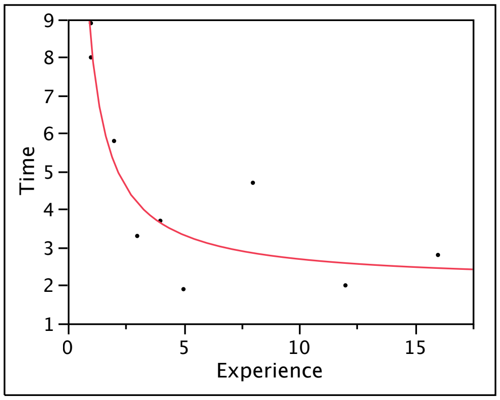
(e) Not assigned. Use the values from the second row of the table shown with part "a". The predicted value is about 848,000.

3.30

(a) See the Anova table in the JMP output. F = 106.303.

(b) The p-value of the F-statistic, like that for the slope, is smaller than any reasonable α. Reject H_0 and conclude there is statistically significant linear association.

(c) As in "b".



Transformed Fit to Reciprocal

Time = 2.0575254 + 6.3537457*Recip(Experience)

Summary of Fit

RSquare	0.855361
RSquare Adj	0.834698
Root Mean Square Error	1.029361
Mean of Response	4.566667
Observations (or Sum Wgts)	9

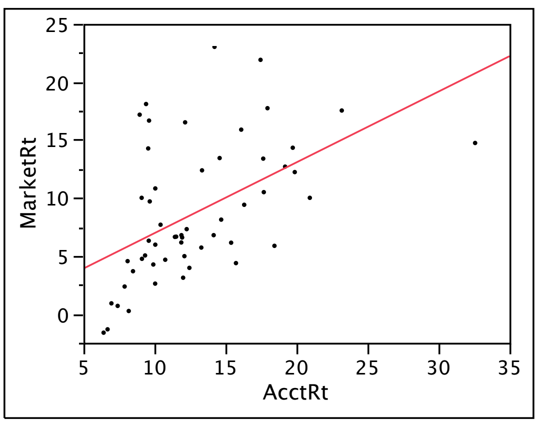
Parameter Estimates

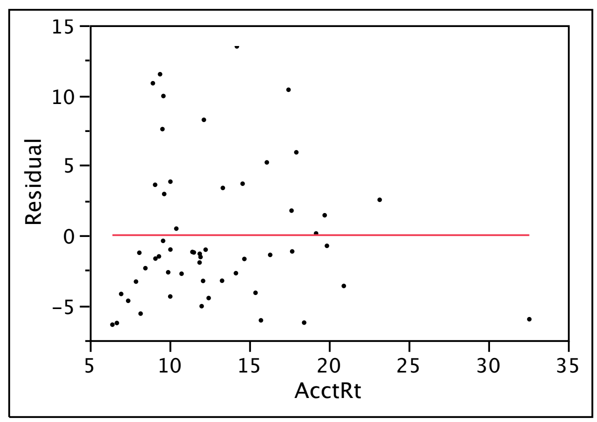
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.0575254	0.519439	3.96	0.0055*
Recip(Experience)	6.3537457	0.987527	6.43	0.0004*

(d)As in "b".

(e) F = 106.303. sqrt(106.303) ≈ 10.31, which is the t-statistic for the slope. The corresponding critical values also match up, namely $(t_{0.025,28} = 2.0484)^2 = 4.196$ and $F_{0.05,1,28} = 4.196$.

3.35 The following figure and output shown above summarize the regression on the reciprocal (Note that there's a point is misplaced in Figure 3.31 of the text.)





Summary of Fit

RSquare	0.259349
RSquare Adj	0.245106
Root Mean Square Error	5.084542
Mean of Response	8.740926
Observations (or Sum Wgts)	54

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.8468477	1.975128	0.43	0.6699
AcctRt	0.6104893	0.143067	4.27	<.0001*

In the figure, you can see the estimates for the intercept (2.06) and slope (6.35) noted in the text when Time is regressed on the reciprocal of experience. The prediction then

$$\hat{y} = 2.0575 + 6.3537/5 \approx 3.328$$

3.36 It's odd when an author suggests that you fit a linear regression to data that don't match up very well to the assumptions of the SRM. The following plot graphs market return on accounting rate. The variance of the data around the fit at the far left does not seem constant, the residuals are not symmetric, and there's a leverage point at the far right (American Home Products).

(a) The point estimate of the market rate is $0.8468 + 0.6105 * 15 \approx 10.0$. The standard error of the prediction as an estimate of the mean return (via JMP) is about 0.7526 so that the 95% confidence interval for the mean is ($t_{0.025,52} = 2.007$)

$$10.0 \pm 2.007 * 0.7526 \approx 8.49 \text{ to } 11.51$$

(b) For the point prediction, the standard error jumps up (via JMP) to 5.140 so that the 95% prediction interval is the much wider range given by

$$10.0 \pm 2.007 * 5.14 \approx -0.32 \text{ to } 20.32$$