Answers to Exercises: Module 1

1. The following tables are from the Witness data set.

Descriptive Statistics

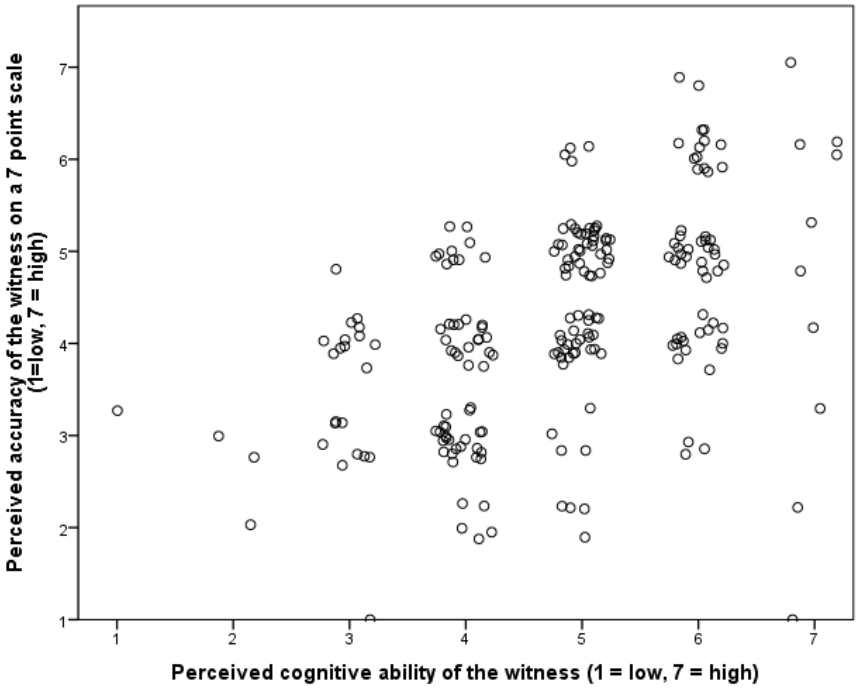
	N	Minimum	Maximum	Mean	Std. Deviation
Perceived accuracy of the witness on a 7 point scale	220	1	7	4.22	1.105
Perceived cognitive ability of the witness (1 = low, 7 = high)	220	1	7	4.83	1.113
Perceived honesty of the witness on a 7 point scale (1 = low, 7 = high)	220	2	7	5.91	.920
Perceived memory of the witness on a 7 point scale (1 = low, 7 = high)	220	2	7	4.45	1.052
Amount of time participant spends with seniors at home (10 = highest)	218	1	10	5.52	2.769
Amount of time participant spends with seniors at school (10 = highest)	216	1	10	3.79	2.643
Valid N (listwise)	214				

Pearson Correlations

	Perceived accuracy of the witness on a 7 point scale	The age of the witness in the fictitious trial	Perceived cognitive ability of the witness	Perceived honesty of the witness on a 7 point scale	Perceived memory of the witness on a 7 point scale	Amount of time participant spends with seniors at home	Amount of time participant spends with seniors at school
Perceived accuracy of the witness on a 7 point scale	1.000	.137*	.469**	.396**	.580**	-.076	.053
The age of the witness in the fictitious trial (49, 69, 79, 89)	.137*	1.000	.027	.162*	-.004	-.068	.021
Perceived cognitive ability of the witness	.469**	.027	1.000	.369**	.561**	.001	-.013
Perceived honesty of the witness on a 7 point scale	.396**	.162*	.369**	1.000	.380**	-.139*	-.074
Perceived memory of the witness on a 7 point scale	.580**	-.004	.561**	.380**	1.000	-.001	-.008
Amount of time participant spends with seniors at home	-.076	-.068	.001	-.139*	-.001	1.000	.500**
Amount of time participant spends with seniors at school	.053	.021	-.013	-.074	-.008	.500**	1.000

Here is a scatter plot between Cognitiv (X) and Accurate (Y). Please note that both variables are discrete, but I used the "jitter" function. That means that it adds a small random component to each point so that all points with the same value on both X and Y are not on top of each other, and you can therefore see how many there are at each point.

Describe the relationship you see, if any, between perceived accuracy and perceived cognitive ability of the witness.



The graph shows a clear positive relationship between perceived cognitive ability and perceived accuracy. This corresponds to the .469 correlation between these variables.

Below are the results from a simple linear regression regressing Accurate (Y) on Cognitiv (X). What are the intercept and slope? What do they tell you about perceptions of Cognitive ability and Accuracy of witnesses?

Regression Coefficients

Dependent Variable:Perceived accuracy of the witness on a 7 point scale

Variable	B	se	t	P	95% Confidence Interval	
Intercept	1.972	.294	6.703	.000	1.392	2.552
COGNITIV	.465	.059	7.833	.000	.348	.582

Intercept is 1.972
Slope is .465

The intercept doesn't tell much, since Cognitiv is on a 1-7 scale. The Intercept gives the average value of perceived accuracy when perceived cognitive ability = 0. Since perceived cognitive ability never = 0, the intercept is meaningless. It is useful for calculating predicted values, though.

The slope, at .465, says that for each 1-unit difference in perceived cognitive ability, the average difference is perceived accuracy increases by about half a unit.

Below are results from a multiple linear regression regressing Accurate on all the numerical predictors.

Regression Coefficients

Dependent Variable:Perceived accuracy of the witness on a 7 point scale

Variable	B	se	t	p	95% Confidence Interval	
Intercept	-.065	.500	-.129	.897	-1.051	.922
AGE_WITN	.006	.004	1.546	.124	-.002	.015
COGNITIV	.174	.064	2.717	.007	.048	.300
HONEST	.156	.071	2.191	.030	.016	.296
MEMORY	.480	.069	6.980	.000	.344	.616
SENIOR_H	-.046	.025	-1.867	.063	-.094	.003
SENIOR_S	.053	.025	2.067	.040	.002	.103

What is the intercept? What does it tell you?

The intercept is now -.065. This is the mean of Perceived Accuracy when all predictors = 0. Once again, a number of predictors never = 0, so it doesn't mean much.

What is the coefficient for each of the predictors? For each, write a single sentence that explains what it means.

Age of Witness: .006 — For each one year difference in age of the witness, the perceived accuracy of the witness increases .006. This is not significantly different from 0. Age of Witness has no apparent effect on perceived accuracy, above and beyond the effects of the other predictors.

Cognitiv: .174 — For each one unit difference in perceived cognitive ability of the witness, the average perceived accuracy of the witness increases .174, after controlling for other predictors. This is significantly different than 0.

Honest: .156 — For each one unit difference in perceived honesty of the witness, the average perceived accuracy of the witness increases .156, after controlling for other predictors. This is significantly different than 0.

Memory: .480 — For each one unit difference in perceived memory of the witness, the average perceived accuracy of the witness increases by .480, almost half a point, after controlling for other predictors. This is significantly different than 0.

Senior_H: -.046 — After accounting for other predictors in the model, the perceived accuracy of the witness decreases by .046 points for each one unit increase in the amount of time a participant spends with seniors at home. This effect is only marginally significant, however, so the effect is not likely different from horizontal.

Senior_S: .053 — For each one unit difference in the time participants spend with seniors, at school, however, there is a significant .053 positive difference in their perception of the witness's accuracy. This is true above and beyond the effects of Witness Age, Perceived Cognitive ability, Perceived Honesty, Perceived Memory, and the amount of time they spend with seniors at home.

Is the coefficient the same for Cognitiv as it was in the bivariate regression? What does that tell you about the effect of perceived cognitive ability on perceived accuracy of eyewitnesses?

No. It is much smaller. There must be a great deal of overlap between Perceived Cognition and other predictors in the variation in Perceived Accuracy they explain. A look back at the correlation table shows that Perceived Cognition has a moderately high correlation with both Perceived Honesty and Perceived Memory.

So the effect of Perceived cognitive ability on perceived accuracy is above and beyond the effect of the other variables in the model.

2. The following tables are from the Births data set. Use all of them to help you answer the questions.

Recall:
bwt_pnds = Birth weight in pounds
pre_wgt = Maternal prepregnancy weight
wtgain = Maternal weight gain
del_wgt = Maternal delivery weight

Descriptive Statistics

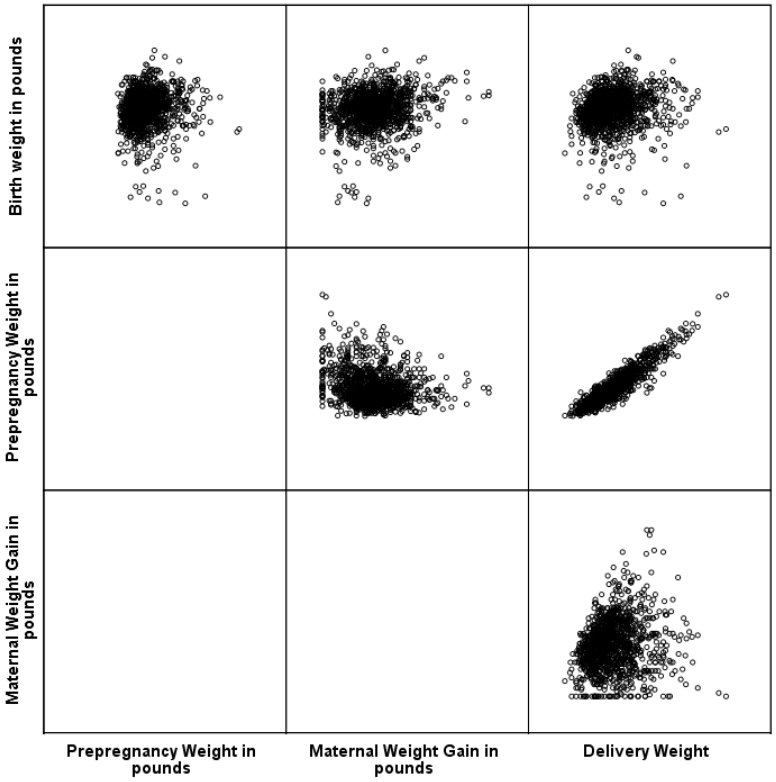
	N	Minimum	Maximum	Mean	Std. Deviation
bwt_pnds	1200	.56	11.38	7.2208	1.29469
pre_wgt	1200	89	375	155.54	39.917
wtgain	1200	0	98	30.24	14.907
del_wgt	1200	100	385	185.72	39.403

Correlations

	bwt_pnds	wtgain	pre_wgt	del_wgt
bwt_pnds	1	.185**	.063*	.134**
wtgain	.185**	1	-.224**	.156**
pre_wgt	.063*	-.224**	1	.928**
del_wgt	.134**	.156**	.928**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).



Below are the results from a simple linear regression regressing infant birthweight (Y) on *maternal prepregnancy weight* (X). What are the intercept and slope? What do they tell you about the predictive relationship between maternal prepregnancy weight and infant birthweight?

Regression Coefficients

Dependent Variable: bwt_pnds

Variable	B	se	t	p	95% Confidence Interval	
Intercept	6.901	.150	45.954	.000	6.606	7.196
pre_wgt	.002	.001	2.198	.028	.000	.004

In the following tables, we have run two more regression models.

How does the intercept change across models? What does it tell you?

The intercept gets slightly smaller, but it still doesn't mean much. It is the mean birth weight when all predictors in the model = 0. As the only one that can possibly equal 0 is weight gain, the intercepts themselves don't mean anything.

The second adds *weight gain during pregnancy* to the model, and the third also adds *maternal weight at delivery*. All variables are measured in pounds.

What is the coefficient for each of the predictors? For each, write a single sentence that explains what it means, looking at all three models.

Regression Coefficients

Dependent Variable: bwt_pnds

Variable	B	se	t	p	95% Confidence Interval	
Intercept	6.115	.183	33.434	.000	5.756	6.474
pre_wgt	.004	.001	3.803	.000	.002	.005
wtgain	.018	.003	7.227	.000	.013	.023

The coefficient for prepregnancy weight is .004. This indicates that after accounting for the effect of weight gain, for each one pound increase in the mother's prepregnancy weight we see an average difference in the infant's weight of .004 pounds.

The coefficient for weight gain is much larger. After accounting for prepregnancy weight, for each one pound difference in weight gain over the pregnancy, we see an average difference in the infant's weight of .018 pounds. In other words, on average, gaining more weight during a pregnancy is associated with slightly bigger babies, on average. This is above and beyond the effect of the size of the mother before pregnancy.

ANOVA

Dependent Variable: bwt_pnds

Source	Type III Sum of Squares	df	Mean Square	F	p
Model	92.827[a]	3	30.942	19.305	.000
pre_wgt	1.309	1	1.309	.817	.366
wtgain	2.503	1	2.503	1.562	.212
del_wgt	1.072	1	1.072	.669	.414
Error	1916.966	1196	1.603		
Corrected Total	2009.793	1199			

a. R Squared = .046 (Adjusted R Squared = .044)

Regression Coefficients

Dependent Variable: bwt_pnds

Variable	B	se	t	p	95% Confidence Interval	
Intercept	6.102	.184	33.226	.000	5.742	6.462
pre_wgt	.037	.041	.904	.366	-.044	.119
wtgain	.052	.042	1.250	.212	-.030	.135
del_wgt	-.034	.041	-.818	.414	-.115	.047

Here we see some very interesting results. Even though the model as a whole is significant, none of the predictors are.

Furthermore, the regression coefficients for both predictors have changed a lot from the previous model and the standard errors are all high compared to the size of the coefficients. Here we see a situation where the correlations among the predictors are all high enough that we've got at least moderate multicollinearity going on. The correlation coefficient between prepregnancy weight and delivery weight is over .90. So delivery weight seems to be confounded with the combination of prepregnancy weight and weight gain and doesn't need to be in the model.