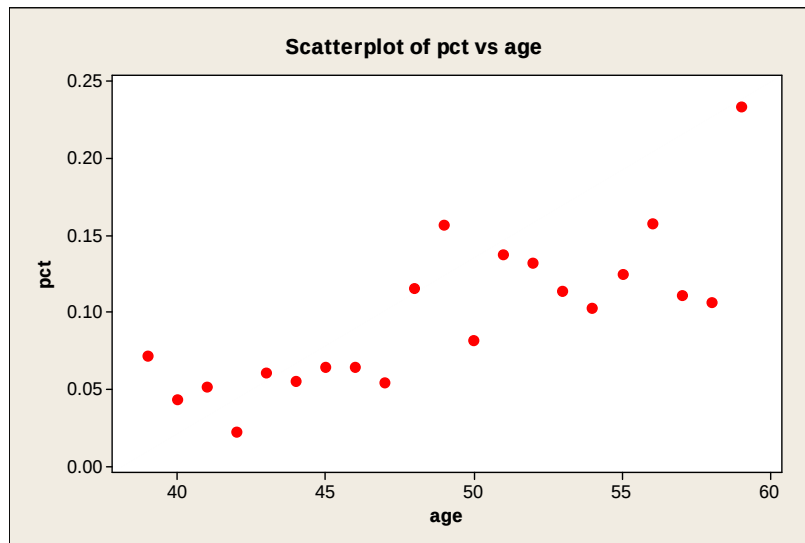
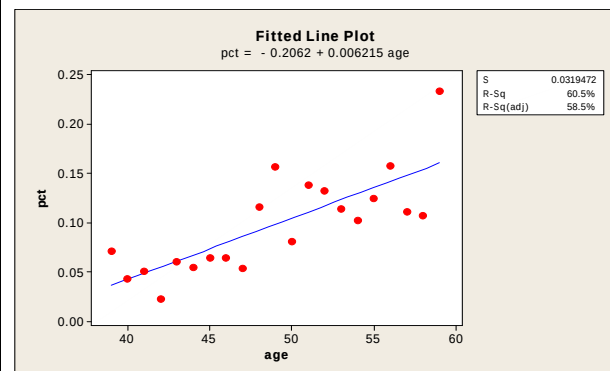


WCGS: CHD as a function of age (text, p.158ff)**Regression Analysis: pct versus age**

The regression equation is
 $pct = -0.206 + 0.00622 \text{ age}$

Predictor	Coef	SE Coef	T	P
Constant	-0.20620	0.05684	-3.63	0.002
age	0.00622	0.00115	5.40	0.000

S = 0.0319472 R-Sq = 60.5%
 R-Sq(adj) = 58.5%

**Binary Logistic Regression: yy, num versus age**

Link Function: Logit
 Response Information

Variable	Value	Count
yy	Event	257
	Non-event	2897
num	Total	3154

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
Constant	-5.93952	0.549322	-10.81	0.000		Lower	Upper
age	0.0744226	0.0113024	6.58	0.000	1.08	1.05	1.10

Log-Likelihood = -869.178

Test that all slopes are zero: G = 42.888, DF = 1, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	26.3069	19	0.122
Deviance	26.1205	19	0.127
Hosmer-Lemeshow	8.2433	6	0.221

ECMO Example Revisited

How shall we analyze these data using Logistic Regression methods?

	Baby Survives	Baby Dies	Total
ECMO	18	2	20
CMT	19	11	30

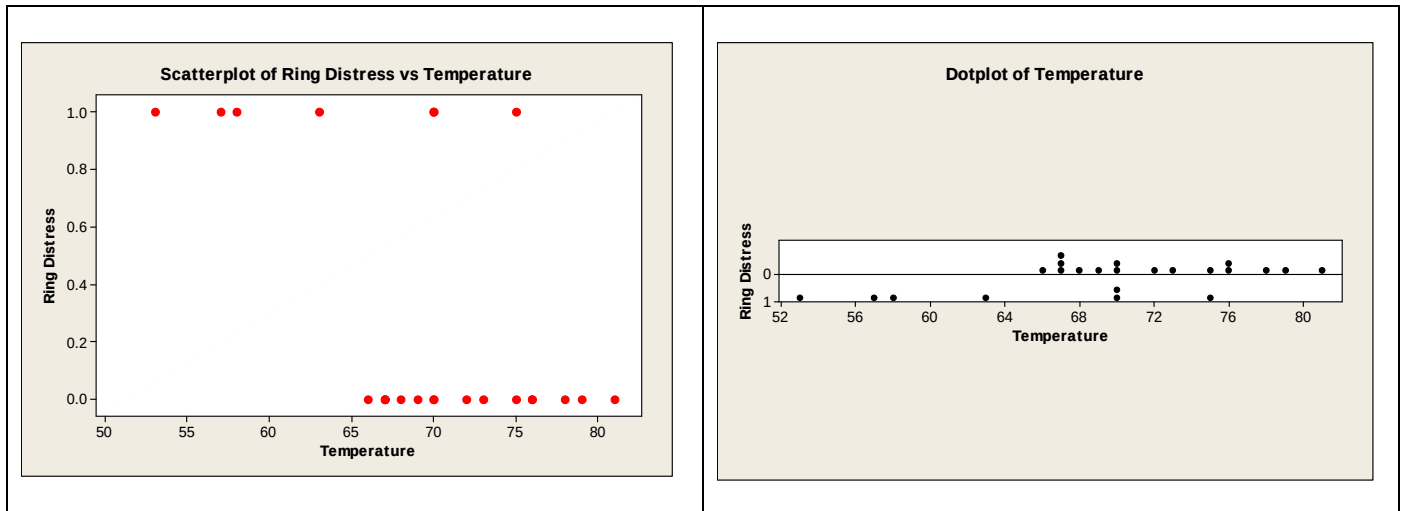
Write down the (logistic) model here.

Interpret the slope parameter estimate here.

Does the LD_{50} estimate have any interpretation here? Why/why not?

Challenger Space Shuttle, O-Ring Failures, and Temperature

References: (1) Hogg & Tanis, Probability and Statistical Inference, 2010, p.328, (2) Agresti, An Introduction to Categorical Data Analysis, 2007, pp. 123-4. **The actual temperature on 1/28/86 was 31F.**



SLR = Simple Linear Regression

Regression Analysis: Ring Distress versus Temperature

The regression equation is

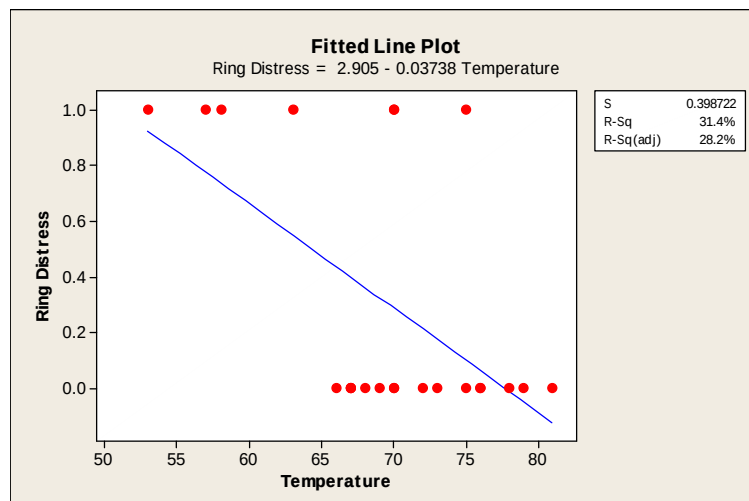
$$\text{Ring Distress} = 2.90 - 0.0374 \text{ Temperature}$$

Predictor	Coef	SE Coef	T	P
Constant	2.9048	0.8421	3.45	0.002
Temperature	-0.03738	0.01205	-3.10	0.005

S = 0.398722 R-Sq = 31.4% R-Sq(adj) = 28.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1.5310	1.5310	9.63	0.005
Residual Error	21	3.3386	0.1590		
Total	22	4.8696			



Binary Logistic Regression**Binary Logistic Regression: Ring Distress versus Temperature**

Link Function: Logit

Response Information

Variable	Value	Count
Ring Distress	1	7 (Event)
	0	16
	Total	23

Logistic Regression Table

Odds	95% CI	Coef	SE Coef	Z	P	Ratio	Lower	Upper
Predictor								
Constant		15.0429	7.37862	2.04	0.041			
Temperature		-0.232163	0.108236	-2.14	0.032	0.79	0.64	0.98

Log-Likelihood = -10.158

Test that all slopes are zero: G = 7.952, DF = 1, P-Value = 0.005

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	11.1303	14	0.676
Deviance	11.9974	14	0.607
Hosmer-Lemeshow	9.7119	8	0.286

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	85	75.9	Somers' D 0.56
Discordant	22	19.6	Goodman-Kruskal Gamma 0.59
Ties	5	4.5	Kendall's Tau-a 0.25
Total	112	100.0	

How do we interpret the slope parameter estimates?A. In Simple Linear Regression:B. In Binary Logistic Regression:C. What about LD50?

Tobacco Budworms (Collett, 2003)

		Number (and Percent) of Insects Dying of 20 Insects					
	Dose	1.0	2.0	4.0	8.0	16.0	32.0
	$x = \log_2(\text{Dose})$	0	1	2	3	4	5
Insect	Female	0 (0%)	2 (10%)	6 (30%)	10 (50%)	12 (60%)	16 (80%)
Sex	Male	1 (5%)	4 (20%)	9 (45%)	13 (65%)	18 (90%)	20 (100%)

**Binary Logistic Regression:
Y_FEMALE, n versus DOSE**

Link Function: Logit

Response Information

Variable	Value	Count
Y_FEMALE	Event	46
	Non-event	74
n	Total	120

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P
Constant	-1.71578	0.322323	-5.32	0.000
DOSE	0.115676	0.0237872	4.86	0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	10.5204	4	0.033
Deviance	13.5302	4	0.009
Hosmer-Lemeshow	10.5204	4	0.033

**Binary Logistic Regression:
Y_MALE, n versus DOSE**

Link Function: Logit

Response Information

Variable	Value	Count
Y_MALE	Event	65
	Non-event	55
n	Total	120

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P
Constant	-1.92771	0.401955	-4.80	0.000
DOSE	0.297234	0.0625451	4.75	0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	4.24797	4	0.373
Deviance	4.63398	4	0.327
Hosmer-Lemeshow	4.24797	4	0.373

**Binary Logistic Regression:
Y_FEMALE, n versus LOG2_DOSE**

Link Function: Logit

Response Information

Variable	Value	Count
Y_FEMALE	Event	46
	Non-event	74
n	Total	120

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P
Constant	-2.99354	0.552699	-5.42	0.000
LOG2_DOSE	0.906036	0.167101	5.42	0.000

Log-Likelihood = -57.607

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	2.19580	4	0.700
Deviance	3.11276	4	0.539
Hosmer-Lemeshow	2.19580	4	0.700

**Binary Logistic Regression:
Y_MALE, n versus LOG2_DOSE**

Link Function: Logit

Response Information

Variable	Value	Count
Y_MALE	Event	65
	Non-event	55
n	Total	120

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P
Constant	-2.81855	0.547955	-5.14	0.000
LOG2_DOSE	1.25895	0.212050	5.94	0.000

Log-Likelihood = -48.132

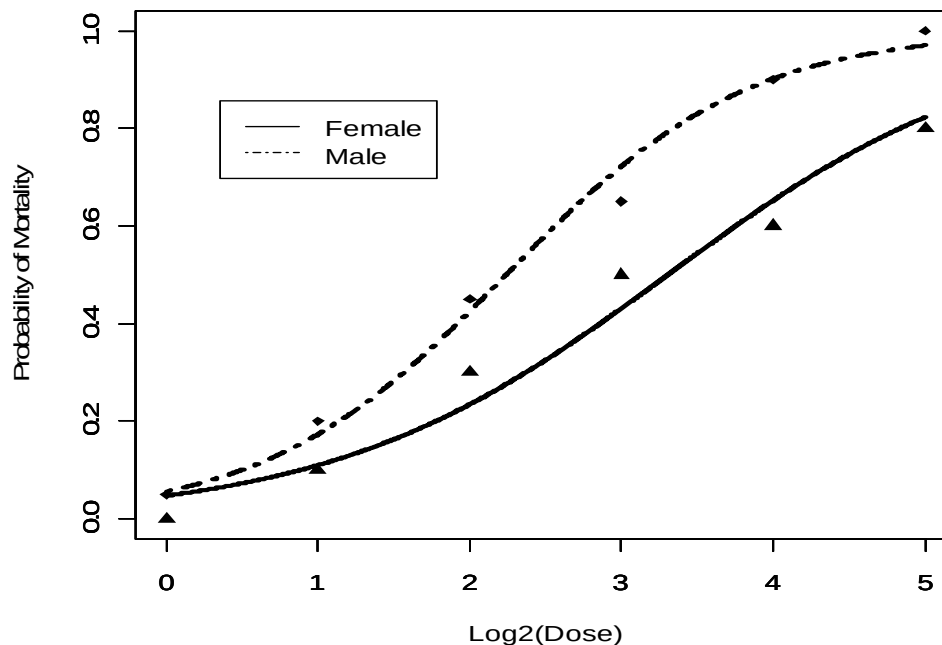
Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1.30890	4	0.860
Deviance	1.88097	4	0.758
Hosmer-Lemeshow	1.30890	4	0.860

Conclusion thus far:

	Parameter Estimates (Standard Errors)		
Sex	b_0 (intercept)	b_1 (slope)	LD ₅₀ on Dose Scale
Females	-2.9935 (0.5527)	0.9060 (0.1671)	9.877 (???)
Males	-2.8186 (0.5480)	1.2589 (0.2120)	4.720 (???)

Do the intercept and slope estimates look similar across the sexes? Can one curve be used for both sexes?



Binary Logistic Regression:
Num_Dead, Num_Exposed versus L2D, FEM, ...

Response Information

Variable	Value	Count
Num_Dead	Event	111
	Non-event	129
Num_Exposed	Total	240

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P
Constant	-2.81855	0.547987	-5.14	0.000
L2D	1.25895	0.212065	5.94	0.000
FEM	-0.174987	0.778310	-0.22	0.822
L2D*FEM	-0.352913	0.269990	-1.31	0.191

Log-Likelihood = -105.739

Test that all slopes are zero: G = 119.882, DF = 3, P-Value = 0.000

Binary Logistic Regression:
Num_Dead, Num_Exposed versus L2D

Response Information

Variable	Value	Count
Num_Dead	Event	111
	Non-event	129
Num_Exposed	Total	240

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P
Constant	-2.76609	0.370134	-7.47	0.000
L2D	1.00681	0.123589	8.15	0.000

Log-Likelihood = -111.734

Test that all slopes are zero: G = 107.892, DF = 1, P-Value = 0.000

Write down the Full Model Function, Null and Alternative Hypotheses, TS, p-value and your clear conclusion.