Additional Exercises

Chapter 1

(none)

Chapter 2

AE 2.1 Here are two probability tables. In each table compute P(A|B), P(A), and P(B).

Table I	B^C	B
A^C	0.20	0.05
A	0.20	0.55

Table II	B^C	B
A^C	0.06	0.14
A	0.24	0.56

In Tables I and II which of the following is true?

a) P(A | B) > P(A) b) P(A | B) < P(A) c) P(A | B) = P(A)

Which condition (a, b, or c) implies that A and B are independent?

AE 2.2 An old deck of cards is missing a few. Here's what's left:

BLACK	**Spades:**	A	2	3	4	5	6	7	8	J	K	
	Clubs:	2	3	4	5	6	7	8	Q	K		
RED	**Hearts:**	A	4	5	7	8	9	J	Q	K		
	Diamonds:	A	2	3	4	5	6	7	9	10	J	K

Face Cards

One card is selected at random from a well shuffled deck. F = Face, R=Red

a) Compute $P(F \cup R)$, *and* $P(F \mid R^C)$.

b) Are F and R independent? (Justify your answer,)

Chapter 3

AE 3.1 Suppose that P(A) =.3 and P(B) = .5, and P(A ∩ B) = .20.

	A	A^C	
B			
B^C			

a) Fill in the joint and marginal probabilities in the table.

b) Are A and B independent? No ☐ Yes ☐ Why? ____

____

AE 3.2 Mom had the frugality of most people who grew up in the great depression. For years she used an old dented tin measuring cup. I told my dad that it wouldn't measure properly, but he solemnly assured me that the dent wouldn't change the capacity of the cup as long as the metal was not broken. I believed him for a long time until one day I pictured in my mind what would happen if I totally flattened the cup – it would hold next to nothing.

What is the name of this type of reasoning? What is the purpose?

a) Name: ____

b) Purpose: ____

____

AE 3.3 What is the operational definition of my subjective probability for the sentence, "A woman will be elected president of the U.S. before 2025."

____

____

AE 3.4 At one point during the 2000 presidential election, the Iowa Political Market website (www.biz.uiowa.edu/iem/) the asking price was $0.64 for a futures contract that paid $1 if Al Gore won the election and $0.38 for a contract that paid $1 if George Bush won. Suppose you bought both contracts . Why would that have been irrational? What is the name of the type of financial situation that you would have placed yourself?

a) Why irrational? ____

____

b) Name? ____

AE 3.5 You are one unlucky person. On any given weekday you think that there is probability 0.3 that you will oversleep and probability 0.5 that you will be stuck in traffic on your way to work. Assuming these events are independent what is your probability that you'll oversleep AND get stuck in traffic on the same day?

AE 3.6 Sally habitually bets on Red at roulette. Although the house odds (1 to 1) are not fair and she can expect to lose money in the long run this is not a Dutch book. Explain why.

AE 3.7 In the next presidential election, Bill's fair price for "The Republican candidate will win 250 or more electoral votes," is \$0.70 and his fair price for "The Republican candidate will win 300 or more electoral votes," is \$0.53.

Assuming Bill's prices are rational

a) Why must the price on the first sentence larger than the price of the second sentence.

b) What is Bill's fair price for the sentence, "Bush will win 251 to 300 electoral votes."

AE 3.8 Jill is playing roulette (for the purposes of this exercise assume that she believes that each of the 38 numbers is equally probable). On a whim, Jill puts \$1 on Red and \$4 on Even. What is her expected value? Is this a Dutch Book?

Chapter 4

AE 4.1 Make a stemplot (stem-and-leaf diagram) of the following observations (do not split the stems). Compute the five number summary (use the instructor's definition of quartiles).

3.5 5.7 6.3 5.1 5.5 6.2 5.2 4.4 3.5 3.9 7.1 1.5
5.5 2.8 7.6 6.1 3.3 2.9 4.4 5.0 4.1 6.6 3.6

Stems	Leaves

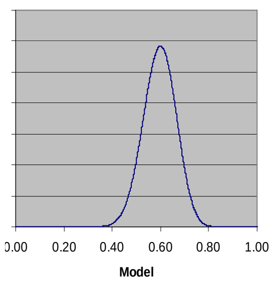
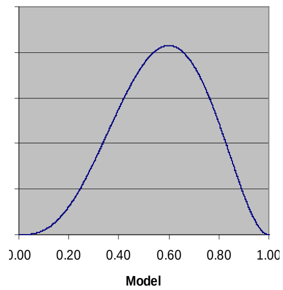
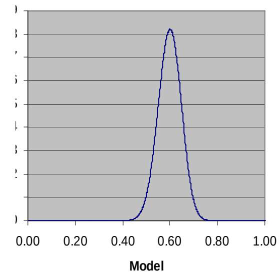
5 Number summary

AE 4.2 Compute the mean and standard deviation of these numbers: 63, 48, 55, 62, 52, 44.

$\bar{x} =$	
$s =$	

Chapter 5

AE 5.1 Here are three posterior distributions for the proportion (p) of black chips in a box. These are based on different n's. Arrange this sequence of pictures in increasing order from the smallest n to the largest. Explain your reasoning.

A	B	C
		

a) Smallest n (enter A, B, or C):_____, Next _____, Largest ______.

b) Explanation: ____

____

AE 5.2 There are three cups. A, B, and C. Cup A contains 3 B and 7 W chips. Cup B contains 5 B and 5 W chips. Cup C contains 8 B and 2 W chip. Someone selects a cup at random and hands it to you. You can't look in the cup, but you can make as many draws with replacement as you like. You made n = 5 draws and got 3 W and 2 B chips. Compute the posterior probabilities of the three models (A, B, or C) given these data. Label the columns.

Model				
A (3B 7W)				
B (5B 5W)				
C (8B 2W)				

AE 5.3 MnSOST-R is a diagnostic "test" to predict if a paroled sex offender will recidivate; i.e., commit another violent sexual offense (V) sometime in the future. The sensitivity of the test is P(+ | V) = 0.17 and the specificity is P(- | V^c) = 0.97 . Assuming that the prevalence of violent sexual recidivism is P(V) = 0.30, use Bayes rule (or the flow chart method in chapter 1) to compute P(V | +) and P(V^c | –).

a) P(V | +) = ____

b) P(V^c | –) = ____

AE 5.4 A box model contains 700 black and 300 white chips. You select n=5 chips independently with replacement and obtained BBBWB in exactly that order. What is the probability of these data given the model? What is the technical term for the probability of the data given a model.

a) P(BBBWB | Model) = ____

b) This is called the ____

Chapter 6

AE 6.1 A sample of n=200 university students contains 12% left-handers. Obtain a 99% ***(not the usual 95%)*** credible interval for the true percent of left-handers among all UI students.

AE 6.2 A sample of n=250 drawn from a population of people over the age of 65 contained x = 50 people who are still working. Compute $\hat{p}$ and SEP and state the approximate posterior distribution of the proportion (p) of workers in the population. Obtain an approximate 95% credible interval for p.

a) The posterior distribution is: ______________________________

____

b) 95% Credible interval: ______________ to ______________

AE 6.3 I took a sample of n=200 bolts made to attach bumpers to sport utility vehicles. The sample contained 15 defectives. Describe the posterior distribution of the proportion p of defectives in the population. Justify your answer. Obtain a 95% credible interval for the population proportion.

a) The posterior distribution is ______________________________

b) The 95% credible interval is ______________ $\leq p \leq$ ______________

AE 6.4 I took a sample of n=20 cardiac pacemakers and found that 2 of them had cracked insulation in the lead wires. Why can I not compute the 95% credible interval using the same method as in question 1?

____

____

____

AE 6.5 Suppose that the posterior distribution of p is normal with $\mu = 0.78$ and $\sigma = 0.13$. Compute the probability that p is bigger than 0.5 . Compute the probability that p is between 0.6 and 0.9 .

a) $P(p > .5) =$ ______________________________

b) $P(0.6 < p < 0.9) =$ ______________________________

AE 6.6 A sample of n=400 drawn from a population of people over the age of 65 contained 30 people who were still working. Describe the approximate posterior distribution of the proportion (p) of workers in the population.

Shape:	a) Normal	b) t(29)	c) Highly Skewed		
$\mu =$	a) 0.013	b) 0.075	c) 0.30	d) 30	e) None of these
$\sigma =$	a) 0.00017	b) 0.013	c) .048	c) 0.075	d) $\sqrt{30}$

AE 6.7 Suppose that **X** has a normal posterior distribution with μ =120 and σ = 15 . Compute P(**X** > 110) and P(105 < X ≤ 125).

a) 95% CI:____________________ to ____________________

b) P(**X** > 110): ____________________

c) P(105 < X ≤ 125):____________________

____

Chapter 7

AE 7.1 Baldus, Pulaski, and Woodworth looked at Georgia felony murder convictions involving armed robbery. The table shows verdicts in cases with a black defendant broken down by race of victim.

	White Victim	Black Victim
n of cases	160	104
# getting the death penalty	60	15

Is there a "statistically significant" difference ($\Delta = p_{white} - p_{black}$) between the proportions of white and black victim cases getting the death penalty? How do you know this?

Compute the posterior probability that the difference, is positive, $P(\Delta > 0 \mid \text{Data})$.

Compute the posterior probability that the difference is between .25 and .30, $P(.25 < \Delta < .30 \mid \text{Data})$.

a) Statistically Significant?: Yes ☐ No ☐ Why?:______________________________

____

b) $P(\Delta > 0 \mid \text{Data}) =$ ______________________________

c) $P(.25 < \Delta < .30 \mid \text{Data}) =$ ______________________________

AE 7.2 Suppose that Δ has an approximately normal posterior distribution with $\mu = 0.358$ and $\sigma = 0.195$. Obtain a 95% credible interval for Δ and compute $P(\Delta > 0 \mid \text{Data})$. Is Δ "significantly different" from 0?

d) 95% CI: ____________________ to ____________________

e) $P(\Delta > 0 \mid \text{Data})$: ______________________________

f) Significant: No ☐ Yes ☐ Why? ______________________________

____

AE 7.3 A study comparing coffee drinkers with non-drinkers reported that the relative risk (RR) of pancreatic cancer for coffee drinkers compared to non-drinkers was not "statistically significant". Here are some hypothetical 95% credible intervals for the Relative Risk (not the difference Δ). Which of these is (are) consistent with the report. Which is (are) consistent with the study report? Which is impossible? Explain why in each case.

a) 95% CI: .85 to 3.24 Consistent: No ☐ Yes ☐ Why? ______________________________

____

b) 95% CI: 1.15 to 3.01 Consistent: No ☐ Yes ☐ Why? ______________________________

____

c) 95% CI: -.53 to .67 Consistent: No ☐ Yes ☐ Why? ______________________________

____

AE 7.4 MacPhail, et al. collected data as part of a study of gender and the risk of HIV infection in a South African township. They reported that, "… by age 20, 43% of females were infected compared to 9% of men." Compute the relative risk and increased risk taking females as the "exposed" group. State the risk factor and the adverse outcome.

a) Relative Risk = ____________________

b) Increased Risk = ____________________

c) Risk Factor: ____________________ Adverse Outcome: ____________________

AE 7.5 A different driving simulator study compared 66 drivers wearing monofocal vs. 43 wearing multifocal lens implants. The table shows (somewhat fictionalized) data on the numbers of subjects (x) who failed to correctly recognize a "Truck Crossing" sign.

	Monofocal	**Multifocal**
n	66	43
x	10	15

State the posterior distribution of $\Delta = p_{Multi} - p_{Mono}$. Obtain a 95% credible interval for **Δ.** Is **Δ** significantly different from 0?

a) Posterior: ____________________

b) 95% CI: ____________________ to ____________________

c) Significant: No ☐ Yes ☐ Why? ____________________

AE 7.6 A prospective study recruited 10,000 smokers and 10,000 non-smokers aged 30 to 39 and followed them for 20 years. The cases of throat cancer are shown in this table.

	no Cancer	Cancer	%
Smokers	9800	200	2.0%
Non-Smokers	9950	50	0.5%

What is the relative risk of throat cancer for smokers vs. non-smokers?

AE 7.7 A 1977 study found that the relative risk of cardiovascular death was 1.15 (95% CI: 1.07 to 1.23) among populations drinking very soft water compared to populations drinking normally hard water. There are about a million cardiovascular deaths in the US every year, so a relative risk of 1.10 or larger would produce tens of thousands of additional deaths a year and would be an important public health problem. What is the probability that the relative risk exceeds 1.00? What is the probability that it exceeds 1.10?

a) P(**RR** > 1 | Data) = ____________________

b) P(**RR** > 1.10 | Data) = ____________________

AE 7.8 In a randomized experiment, 400 children brushed with baking powder and 400 brushed with toothpaste. 52 of the baking powder children (13%) got cavities and 40 of the toothpaste children (10%) got cavities. The difference is 3 percentage points. Obtain a 95% credible interval for the true difference. Is the difference "statistically significant?"

AE 7.9 A random roadside survey of 481 males and 138 females found that 77 males and 16 females had detectable amounts of alcohol by a breathalyzer test. Compute approximate 95% credible intervals for the odds ratio (males/females) and difference (males – females). Is the disparity between male and female rates "statistically significant?"

AE 7.10 A study comparing coffee drinkers with non-drinkers reported that the relative risk (RR) of pancreatic cancer had ***p***-value = 0.15. Which of the following is consistent with that ***p***-value?

a) 95% CI: .85 to 3.24 "Statistically significant"
b) 95% CI: .85 to 3.24 Not "statistically significant"
c) 95% CI: 1.15 to 3.01 "Statistically significant"
d) 95% CI: 1.15 to 3.01 Not "statistically significant"
e) 95% CI: -.53 to .67 Not "statistically significant"

AE 7.11 One possible side effect of antihistamines is drowsiness. In a clinical trial of the drug Suspirizine vs. a placebo, increased drowsiness in the drug group was a side effect. Suppose that the regulatory agency requires a warning label if the relative risk of drowsiness exceeds 1.05 (5% increased risk). The table below lists several hypothetical "scenarios" for the clinical trial.

Scenario	**Relative Risk**	**95% Confidence Interval**
A	1.01	0.99 to 1.03
B	0.99	0.90 to 1.09
C	1.07	1.05 to 1.09

In which scenario(s) is the increased (or decreased) risk significant?

a) A b) B c) C d) A and C e) Other

In which scenario(s) would a warning label not be required?

a) A b) B c) C d) A and B e) Other

In which scenario(s) is the study underpowered?

a) A b) B c) C d) A and B e) Other

AE 7.12 A 1977 study found that the relative risk of cardiovascular death was 1.15 (95% CI: 1.07 to 1.23) among populations drinking very soft water compared to populations drinking normally hard water. There are about a million cardiovascular deaths in the US every year, so a relative risk of 1.05 or larger would produce tens of thousands of additional deaths a year and would be an important public health problem.

What conclusion can be drawn from the 1977 study?

a) Hard water is an important public health problem.
b) Soft water is an important public health problem.
c) Soft water is not a public health problem
d) No conclusion is possible; the study is underpowered.
e) None of these

Suppose the 1977 study had found a relative risk of 1.02 (95% CI 0.91 to 1.14). What conclusion could be drawn in that case?

a) Hard water is an important public health problem.
b) Soft water is an impot public health problem.
c) The RR is insignificartannt; soft water is not a public health problem
d) No conclusion is possible; the study is underpowered.
e) None of these

Chapter 8

AE 8.1 The posterior distribution of Δ is approximately t(11) with $\mu = .37$ and $\sigma = .09$. Obtain a 95% credible interval for Δ. Use tTailArea.xls to compute the probability that Δ is less than 0.50.

98% CI: ____________________ to ____________________

P(Δ < 0.50 | Data): ____________________ to ____________________

AE 8.2 A box model for has box mean μ_{box}. Based on a large sample from the box, the posterior distribution of μ_{box} is approximately normal with $\mu = \bar{x} = 105$ and σ = SEM = 3. Compute $P(\mu_{box} > 100 \mid \text{Data})$, and $P(102 < \mu_{box} < 106 \mid \text{Data})$.

a) $P(\mu_{box} > 100 \mid \text{Data})$ = ____________________

b) $P(102 < \mu_{box} < 106 \mid \text{Data})$ = ____________________

AE 8.3 A clinical trial comparing a new delivery system (treatment) to the standard delivery system (control) for inhaled steroids reported the following statistics for FEV_1 /FVC (a measure of airway obstruction).

	Treatment	Control
n	225	220
$\bar{x}$	63	70
s	25	20

State the posterior distribution of $\Delta = \mu_{Ctrl} - \mu_{Trt}$. Obtain a 95% credible interval for Δ and say if the difference is significant.

a) Posterior distribution: ____________________

b) 95% CI: ____________________ to ____________________

c) Significant: No ☐ Yes ☐ Why? ____________________

d) P(Δ > 0 | Data): ____________________

AE 8.4 A study of the effects of sedating (S) and non-sedating (N) antihistamines on driving impairment was done in the Iowa High Performance Driving Simulator. Statistics for steering instability are shown in the table.

	Sedating	**Non-Sedating**
n	40	40
$\bar{x} \pm$ sem	.53 ± .016	.49 ± .014

State the posterior distribution of $\Delta = \mu_S - \mu_N$ (you do not need to do Satterthwaite's approximation). Obtain a 95% credible interval for Δ. Is the difference significantly different from 0? Why? Compute P(Δ > 0 | Data).

a) Posterior: ____________________

b) 95% CI: ____________________ to ____________________

c) Significant: No ☐ Yes ☐ Why? ____________________

d) P(Δ > 0 | Data) = ____________________

AE 8.5 A study in which healthy men aged 45 to 65 received either a calcium supplement or a placebo for 12 weeks reported, "The calcium group had lower blood pressure (p=.008) compared with the placebo group." (Note: blood pressure is measured in millimeters of mercury, abbreviated mm Hg.)

Which of the following sets of data is consistent with the study report? Why?

A: Difference = 10 mm Hg with 95% credible interval 2.4 to 17.6.
B: Difference = 20 mm Hg with 95% credible interval -5 to 45.

AE 8.6 Two hundred forty-four male alcoholics suffering from secondary hypertension participated in a study to determine the efficacy of a new antihypertensive agent. The men were assigned at random to either the control group or the treatment group. Men in the control group received a placebo. Statistics for arterial pressure at 30 days post treatment for the 87 subjects who completed the study are shown in this Table.

	Arms	
	CTR	TRT
n	100	144
mean ($\bar{x}$)	140.3	102
standard deviation (s)	15.0	7.2

Describe the approximate posterior distribution of the difference (Δ) between the means of the two populations and compute a 95% credible interval for the difference.

The posterior distribution is: ____________________________________

____

____

95% Credible interval: ____________________ to ____________________

AE 8.7 In a study of paranoid schizophrenics, the age of onset of each person's illness was determined. Statistics for age at onset are shown in this table.

Schizophrenia Study	**Females**	**Males**
n	16	9
mean ($\bar{x}$)	29.6	26.5
standard deviation (s)	9.6	9.9

Compute the 95% credible interval for the difference (Δ) between the means of the two populations, carrying all intermediate computations to at least 4 significant figures. Which answer is closest to yours?

CI: a) -26 to 32 b) -6 to 12 d) -9 to 8 c) -5 to 11 e) 5 to 11

What are the degrees of freedom?

df: a) < 9 b) ≈ 9 c) ≈ 16 d) ≈ 25 e) > 25

AE 8.8 Suppose you want to compare two drugs and a placebo (A/B/P). You want each treatment to be administered 36 times. Here are four paragraphs describing ways to assign treatments to subjects.

<table>
<tr><th>A</th><th>B</th></tr>
<tr><td>Recruit 36 subjects.
Prepare a deck of 36 cards this way,
12 cards labeled “A”
12 cards labeled “B”
12 cards labeled “P”
Shuffle the cards and deal them to the subjects.</td><td>Recruit 36 subjects.
Prepare a deck of 36 cards this way,
6 cards labeled “ABP”
6 cards labeled “APB”
6 cards labeled “BAP”
6 cards labeled “BPA”
6 cards labeled “PAB”
6 cards labeled “PBA”
Shuffle the cards and deal them to the subjects.</td></tr>
<tr><th>C</th><th>D</th></tr>
<tr><td>Recruit 216 subjects.
Prepare a deck of 216 cards this way,
36 cards labeled “ABP”
36 cards labeled “APB”
36 cards labeled “BAP”
36 cards labeled “BPA”
36 cards labeled “PAB”
36 cards labeled “PBA”
Shuffle the cards and deal them to the subjects.</td><td>Recruit 108 subjects.
Prepare a deck of 108 cards this way,
36 cards labeled “A”
36 cards labeled “B”
36 cards labeled “P”
Shuffle the cards and deal them to the subjects.</td></tr>
</table>

Which of these describes a crossover design with 36 observations on each treatment?

a) A b) B c) C d) D e) None of these

Which of these describes a completely randomized design with 36 observations on each treatment?

a) A b) B c) C d) D e) None of these

Chapter 9

AE 9.1 Suppose that male and female regression equations are,

Males: $\hat{S} = 106 + 0.7 \cdot B$

Females: $\hat{S} = 100 + 0.6 \cdot B$

The average value of B for males and females combined is 25. Draw a graph of the two equations. Indicate the location of the adjusted mean values of S for males and females. Calculate the adjusted mean difference for Males vs Females and point it out on the graph.

AE 9.2 The patients in the data set listed in the table below are wearing one of two different cochlear implant devices. The x-variable is pre-implant lip reading ability and the y-variable is speech understanding after nine months of experience with the implant. The table below shows four different ways to set up the design variables.

			A			**B**			**C**			**D**		
Device	Lip read (x)	Hear (y)	X1	X2	X3	X1	X2	X3	X1	X2	X3	X1	X2	X3
A	20	31.00	1	20	0	1	1	0	1	0	20	1	20	31.00
A	35	70.33	1	35	0	1	1	0	1	0	35	1	35	70.33
A	26	48.75	1	26	0	1	1	0	1	0	26	1	26	48.75
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
B	0	25.50	1	0	0	1	0	1	0	1	20	1	0	25.50
B	22	44.00	1	0	22	1	0	1	0	1	35	1	22	44.00
B	29	40.00	1	0	29	1	0	1	0	1	26	1	29	40.00
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

Suppose I want to have ***different intercepts*** for the two devices but the ***same slope*** for both. Which set of design variables corresponds to a model in which β_1 is the intercept for device A, β_2 is the intercept for device B, and β_3 is the common slope for the two devices?

a) A b) B c) C d) D e) None of these

Suppose I want to have ***different slopes*** for the two devices but the ***same intercept*** for both. Which set of design variables corresponds to a model in which β_1 is the intercept , β_2 is the slope for device A, and β_3 is the slope for device B?

a) A b) B c) C d) D e) None of these

AE 9.3 A statistician computed a multiple linear regression model for the relationship between salary (y), race, gender, and years of experience. This table shows the estimated regression coefficients, their standard errors, and credible intervals.

			95% Credible Interval	
Variable	$\hat{\beta}$	seb	Lower	Upper
Constant	30,000	5,000	20,000	40,000
Male	400	220	40	760
White	700	450	[redacted]	[redacted]
Years	500	60	300	500

To the nearest $1000, what is the predicted average salary for black males with 5 years of experience.

a) $30,000 b) $33,000 c) $35,000 d) $37,000 e) None of these

How much more does the average white male with the same amount of experience earn?

a) $34,000 b) $400 c) $700 d) $2,500 e) None of these

Is the difference "statistically significant"?

a) No b) Yes c) Cannot Determine

AE 9.4 A study was conducted to determine if maternal smoking influenced the birth weight of infants. Birth weight (pounds) is also influenced by gestational age (number of days between conception and birth). The sample included n = 100 babies. Design variables were the intercept, whether the mother smoked during pregnancy (1=yes/0=no), and the gestational age of the infant (days). The average gestational age was 233 days. The linear (not logistic) regression results were as follows:

			95% CI	
Variable	$\hat{\beta}$	Std Err	Lower	Upper
Intercept	-1.68	1.30	n/a	
Gestational Age	.052	0.022	0.009	0.095
Mother Smoked	-.25	0.09		

Compute the 95% credible interval for the regression coefficient for maternal smoking.

a) -0.43 to -0.07 b) -0.27 to -0.23 c) -0.34 to -0.16 d) None of these

Does gestational age have a "significant" influence on birth weight?

a) No b) Yes c) Impossible to say without a *p*-value.

What is the adjusted mean difference ($\hat{\Delta}_{adj}$) between the birth weights of babies with non-smoking mothers and babies with smoking mothers?

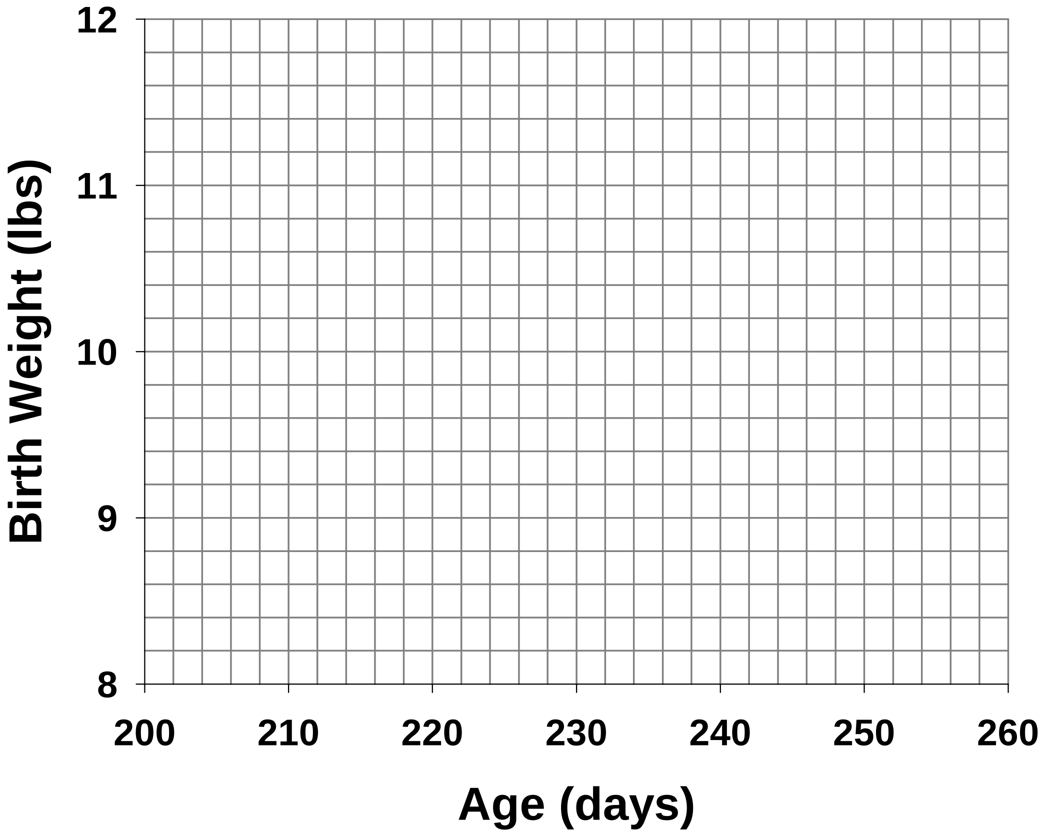
a) 0.052 lb b) 0.09 lb c) 0.25 lb d) 1.68 lb e) 12.2 lb

AE 9.5 Referring to the previous question, compute the credible intervals for the regression coefficients. Does maternal smoking have a significant effect on birth weight? How do you know this?

Use the betas in the table in question 9 to compute predicted values for the four cases shown in the table below. Fill in the numbers in the last column.

Case	Age	Mother Smokes	$\hat{y}$
1	200	No	
2	260	No	
3	200	Yes	
4	260	Yes	

Now use these numbers to draw a graph of the regression model. The y-axis is Birth Weight (vertical) and the x-axis is Gestational Age (horizontal). There will be two upward slanting regression lines -- one for women who smoked during pregnancy and another line for women who did not. The mean age of all babies in the study was 236 days. Label the adjusted mean birth weights for babies whose mothers smoked and did not smoke during pregnancy and draw and label a line segment showing the difference between the adjusted means.



AE 9.6 Referring to the table of betas in AE 9.4. Create a table that shows what the data file looks like. The dependent variable is birth weight and the explanatory variables are mother smoked (Y/N) and the gestational age of the baby. Add three more columns showing the design variables corresponding to the three betas. Suppose that the beta for gestational age was 0.06 for smoking mothers and 0.048 for smoking mothers. Add two more columns to show the design variables corresponding to these betas. Assuming the average gestational age was 236 days. Using the given betas, hand compute the least squares mean birth weights for smoking and for non-smoking mothers and the difference of the least square means (you will not be able to compute the standard errors or credible intervals).

AE 9.7 Replicate the analysis in textbook section 9.7, page 199 (it is not necessary to run the programs on page 200). You should submit a report (not raw computer output) consisting of the following elements: a listing of your program (you may delete parts of the data to make the listing shorter). Paste in enough of the output window to show

1) the date stamp, which will look like this:

```
The SAS System                                         11:52 Thursday, August 17, 2006
```

2) the Solution for Fixed Effects and 3) the Least Squares Means and Differences of Least Squares Means (see page 201)..

Repeat the above using the program on page 202 (you can skip PROC GLMOD)

AE 9.8 Replicate the analysis in textbook section 9.8. Your report should include a listing of your program, the date stamp, and enough output to show the least squares means and differences of least squares means.

AE 9.9 Which SAS System® PROC computes adjusted means, adjusted mean differences, and credible intervals?

a) MIXED **b) LOGISTIC** **c) FREQ** **d) SORT** **e) PRINT**

Chapter 10

AE 10.1 Which SAS System® PROC computes adjusted odds ratios, predicted probabilities, and credible intervals?

a) MIXED **b) LOGISTIC** **c) LIBNAME** **d) SORT** **e) DATA**

AE 10.2 The table below shows the results of a (fictional) study of 10-year survival for men who survived a heart attack. The outcome variable is surviving 10 years after first heart attack (1=survived / 0=died), so the logistic regression model predicts the logarithm of the odds on survival. The explanatory variables are family history of early death (1 = at least one first degree male relative died before the age of 55 / 0 = none), body mass index, and age at the time of the first heart attack.

					95% Confidence Interval for odds ratio	
	Variable	$\hat{\beta}$	SE	odds ratio	Lower	Upper
1	Intercept	8.1548	0.8864	na	na	na
2	Family History	-0.0513	0.0112	[redacted]		
3	Body Mass Index	-0.1020	0.1627	0.903	0.656	1.242
4	Age	-0.1069	0.0220	1.113	1.066	1.162

Table 1. Logistic regression coefficients for 10-year survival of males following first heart attack.

Compute the estimated probability of survival for a 58-year-old with a family history of early death and BMI = 30 .

a) -1.2 b) 0.24 c) 0.31 d) .9996 e) None of these

Which of the explanatory variables is(are) not “statistically significant”?

a) Family History b) Body Mass Index c) Age
d) All are significant e) Two or more are insignificant.

Compute the 95% credible interval for the adjusted odds ratio for patients with a family history of early death vs. patients without a family history of early death.

a) 0.93 to 0.97 b) -0.07 to -0.03 c) -1.0 to 2.9 d) None of these

AE 10.3 Replicate the analysis in textbook section 10.5. Your report should include a listing of your program, the date stamp (see AE 9.7), and enough output to show the beta, adjusted odds ratios, and p-hats.

Chapter 11

AE 11.1 The use of unvented gas cook stoves or water heaters is thought to produce sufficiently high levels of nitrogen dioxide as to cause respiratory problems in children. A meta analysis of NO_2 exposure as a risk factor for respiratory diseases in children under the age of 6 was based on 5 studies listed in Table 1. In most studies the exposed group was children living in homes with gas cook stoves and the unexposed group was children living in homes with electric cook stoves.

Table 1 (Real) Studies of Respiratory Diseases and NO_2 Exposure

Study	Children in	Age (yr)	Risk Factor	Odds Ratio	95% CI
A	28 areas of England and Scotland	6-11	Gas cook stove in home.	1.27	1.16 to 1.40
B	London	< 1	Gas cook stove in home	0.63	0.36 to 1.10
C	6 US Cities	6-10	Gas cook stove in home	1.08	0.96 to 1.37
D	Tayside Scotland	6-10	Unvented water heater in home	1.14	0.86 to 1.50
E	Iowa City, IA	6-12	Gas cook stove in home	1.10	0.79 to 1.53
F	Columbus, OH	6-12	Gas cook stove in home	1.10	0.74 to 1.54

The authors of the meta analysis decided not to use all the studies to reach a consensus. Instead they did one meta-analysis on studies A, C, E, and F and another on studies B and D. They did this because it is clear that studies A – F are not all,

a) Adjusted   b) Exchangeable   c) Independent   d) Significant   e) Retrospective

AE 11.2 Referring to AE 11.1, the consensus odds ratio is 1.18 with 95% credible interval 0.97 to 1.38. Some posterior probabilities are:
$P(OR_{con} > 1.0 \mid data) = 0.96$ and $P(OR_{con} > 1.2 \mid data) = 0.39$.

Is the consensus odds ratio "statistically significant"?

a) No   b) Yes   c) Impossible to say without *p*-value.

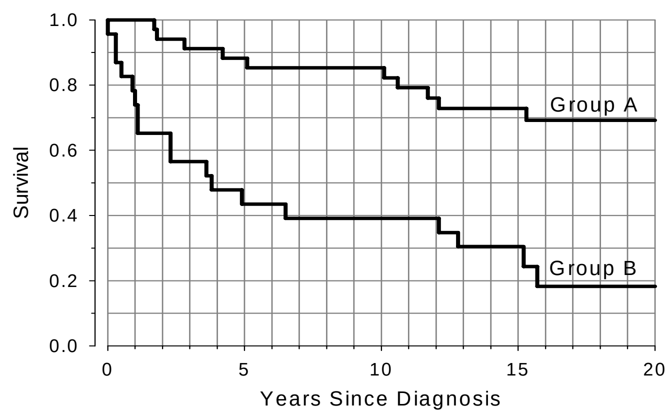
AE 11.3 Referring to AE 11.1, suppose that a consensus odds ratio larger than 1.2 represents an important public health problem. What can you conclude from this meta-analysis?

a) The probability that there is an important public health problem is 0.04
b) The probability that there is an important public health problem is 0.39
c) The probability that there is an important public health problem is 0.61
d) The probability that there is an important public health problem is 0.96
e) None of these.

Chapter 12

AE 12.1 For the coefficients in Table 12.8, compute the survival rates at 12 and 24 months for individuals with this combination of risk factors: performance status 0, Gleason sum <8, LDH=100, alkaline phosphatase=75, no visceral disease, hemoglobin=13.

AE 12.2 Here are Kaplan-Meier survival function estimates for two groups of patients.



In which group does a patient have the greatest chance of survival no matter what the time period.

a) Group A b) Group B c) Inpossible to determine.

What is the median survival time in Group B? (Draw it on the graph on the answer sheet).

a) less than a year b) a little over 2 years c) almost 4 years d) between 6 and 7 years
e) off the chart.

What is the 10-year survival rate for patients in group A? (Draw it on the answer sheet.)

a) 30% b) 40% c) about 44% d) 70% e) 85%

AE 12.3 Download the data file Table12.2.txt from the textbook website and use PROC LIFETEST to compute the Kaplan-Meier survival functions. Your output should include the date stamp, the listings of the three survival curves, the confidence intervals, and the log-rank test.

AE 12.4 Use the 12 and 24 month survival estimates for the reference group described in table 12.7 and the beta values in Table 12.8 to compute the survival rate for patients with performance status 2, Gleason sum 11, LDH 500, alkaline phosphatase 300, PSA 75, no visceral disease, and hemoglobin 13. What are the hazard ratios of this category relative to categories Q1 and Q3? What is the "denominator" category for the raw hazard ratio in equation (12.10)? Is anyone in this category?