

Name: _____ ID _____

There are 14 questions. Mark answers on the answer sheet at the end of the exam and also circle the answer on the exam. You may submit your scratch paper if you expect to ask for partial credit. Tables & formula sheets are distributed separately.

1. 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.

1.1 Shape: a) Normal b) $t(29)$ c) Highly Skewed

1.2 $\mu =$ a) 0.013 b) 0.075 c) 0.30 d) 30 e) None of these

1.3 $\sigma =$ a) 0.000,17 b) 0.013 c) .048 d) 0.075 e) $\sqrt{30}$

2 An unknown quantity, Δ , has a $t(8)$ posterior distribution with $\mu = 0.46$ and $\sigma = 0.22$. Compute the 95% credible interval for Δ .

2.1 a) -0.84 to 1.28 b) -0.05 to 0.97 c) 0.03 to 0.89 d) 0.05 to 0.97 e) None of these

3. Two hundred forty-four subjects participated in a study to determine the efficacy of a new drug. Statistics for reduction in symptoms after treatment are shown in this table. Compute μ and σ of the approximate posterior distribution of the difference (Δ) between the means of the two populations (Treatment – Placebo).

<i>Hypertension Study</i>	Treatment	Placebo
n	100	144
mean (\bar{x})	40.3	2
standard deviation (s)	15.0	7.2

3.1 $\mu = :$ a) 7.8 b) 38.3 c) 40.3 d) 44 e) None of these

3.2 $\sigma = :$ a) 1.6 b) 7.8 c) 16.6 d) $\frac{s}{\sqrt{n}}$ e) None of these

4. 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.

<i>Schizophrenia Study</i>	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 (female – male), carrying all intermediate computations to at least 4 significant figures. Which answer is closest to yours?

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

What are the degrees of freedom?

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

5. A study investigated whether people taking regular doses of aspirin to prevent heart attacks experienced gastric bleeding as a side effect. The data are in this table. Describe the posterior distribution of Δ , the difference between the population rates (users – nonusers).

Aspirin Study	Aspirin Users	Non-Users
n	585	1670
# having heart attacks	24	35

5.1 Distribution: a) nearly normal b) t(12) c) beta(24, 35)

5.2 μ : a) .01 b) .02 c) .04 d) .06 (closest answer to yours)

5.3 σ : a) .00008 b) .008 c) .009 d) .02 (closest answer to yours)

6. One possible side effect of antihistamines is drowsiness. In a clinical trial of the drug Suspirazine 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 1.10 (10% increased risk). The table below lists several hypothetical scenarios for the results of the clinical trial.

Scenario	Relative Risk	95% Credible 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 relative risk statistically significantly increased?

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

In scenario B, compute $P(RR > 1.10 | \text{Data})$. Which answer is closest?

6.2 a) 0.02 b) 0.05 c) 0.5 d) 0.98 e) 2.55

7. An unknown quantity Δ has a t(9) distribution with $\mu = 23$ and $\sigma = 7$.

7.1 Compute the 95% credible interval for Δ .
 7.1 Fill in blanks a) and b) in the ttailarea.xls spreadsheet on the answer sheet to compute $P(\Delta < 15)$. Place an X mark in blank c) or d) (but not both) to show where you expect to find this tail area.
 7.3 Use Table A.2 to get upper and lower bounds on the tail area.

Instructions:			
Enter degrees of freedom and z (or t) value.			
Read left- or right tail area.			
Tail areas for t(df)			
	v =df =	a)	Enter Here
	t=z =	b)	
Left Tail	P(t(df) ≤ z) =	c)	Read Here
Right Tail	P(t(df) > z) =	d)	

8. 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.

Device	Lip read (x)	Hear (y)	A			B			C			D		
			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	0	1	0	25.50
B	22	44.00	1	0	22	1	0	1	0	1	22	1	22	44.00
B	29	40.00	1	0	29	1	0	1	0	1	29	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?

8.1 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?

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

9. A statistician computed a multiple linear (not logistic) 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.

Variable	$\hat{\beta}$	seb	95% Credible Interval	
			Lower	Upper
Constant	30,000	5,000	20,000	40,000
Male	400	220	40	760
White	600	450		
Years	500	60	300	500

To the nearest \$1000, what is the predicted average salary for black males with 6 years of experience.

9.1 a) \$32,000 b) \$33,000 c) \$34,000 d) \$35,000 e) \$37,000

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

9.2 a) \$400 b) \$500 c) \$600 d) \$3,000 e) None of these

Is the difference significant?

9.3 a) No b) Yes c) Cannot Determine

10. 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.

Variable	$\hat{\beta}$	SE	adjusted odds ratio	95% Credible Interval for adjusted odds ratio	
				Lower	Upper
1 Constant	8.1548	0.8864	na	na	na
2 Family History	-0.0513	0.0112	?	?	?
3 Body Mass Index	-0.1020	0.1627	0.903	0.656	1.242
4 Age	-0.1069	0.0220	0.899	0.861	0.938

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

Compute the probability of survival for 64-year-olds with a family history of early death and BMI = 20 .

10.1 a) -0.78 b) 0.46 c) 0.31 d) 0.69 e) 1.00

Which of the explanatory variables is(are) statistically significant?

10.2 a) Family History b) Body Mass Index c) Age
d) None is significant e) Two or more are significant.

Compute the 95% credible interval for the adjusted odds ratio for patients with body mass index 30 vs. patients with body mass index 25. Select the answer closest to yours.

10.3 a) -2.10 to 1.08 b) 0.12 to 2.96 c) 0.47 to 0.73 d) 0.66 to 1.24 e) 0.86 to 0.93

11. 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:

Variable	$\hat{\beta}$	Std Err	95% CI	
			Lower	Upper
Intercept	-1.68	1.30	n/a	
Gestational Age (days)	0.032	.025	-0.017	0.081
Mother Smoked	-0.30	.09	?	?

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

11.1 a) -0.48 to -0.12 b) -0.32 to -0.28 c) -0.39 to -0.21 d) None of these

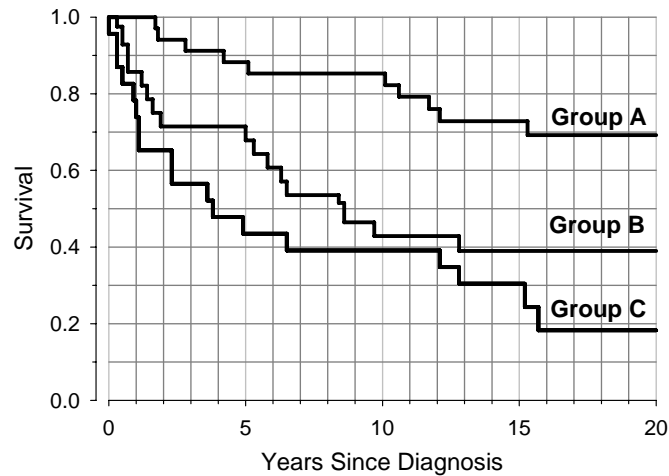
Does gestational age have a significant influence on birth weight?

11.2 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 at the same gestational age with smoking mothers?

11.3 a) 0.032 lb b) 0.025 lb c) 0.09 lb d) 0.30 lb e) 8 lb

12. Here are Kaplan-Meier survival function estimates for three groups of patients.



How do the groups compare in terms of survival time, from longest to shortest?

12.1 a) A>B>C b) C>B>A c) B>A>C d) B>C>A e) Impossible to determine.

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

12.2 a) between 2 and 4 years b) between 4 and 6 years c) between 6 and 8 years
d) between 8 and 10 years e) off the chart.

What is the 10-year survival rate for patients in group C? (Select the closest answer and also draw it on the answer sheet.)

12.3 a) about 30% b) about 40% c) about 50% d) about 60% e) about 70%

13. The next table is the Kaplan-Meier survival function for these survival times:

1 2 5 6 7 9 15 15 >21 21 >25 >30 31 39 40 >40

Interval	at Risk	Failed	Censored	Failure Rate	Survival Rate
0	n/a	n/a	n/a	n/a	1.0000
0 - 1	16	1	0	0.0625	0.9375
1 - 2	15	1	0	0.0667	0.8750
2 - 5	14	1	0	0.0714	C
5 - 6	13	1	0	0.0769	0.7500
6 - 7	12	1	0	0.0833	0.6875
7 - 9	11	1	0	0.0909	0.6250
9 - 15	10	2	0	B	0.5000
15 - 21	A	1	1	0.1250	0.4375
21 - 25	6	0	1	0.0000	0.4375
25 - 30	5	0	1	0.0000	0.4375
30 - 31	4	1	0	0.2500	0.3281
31 - 39	3	1	0	0.3333	0.2188
39 - 40	2	1	1	0.5000	0.1094

13.1 Fill in the missing numbers for cells A, B, and C on the answer sheet.

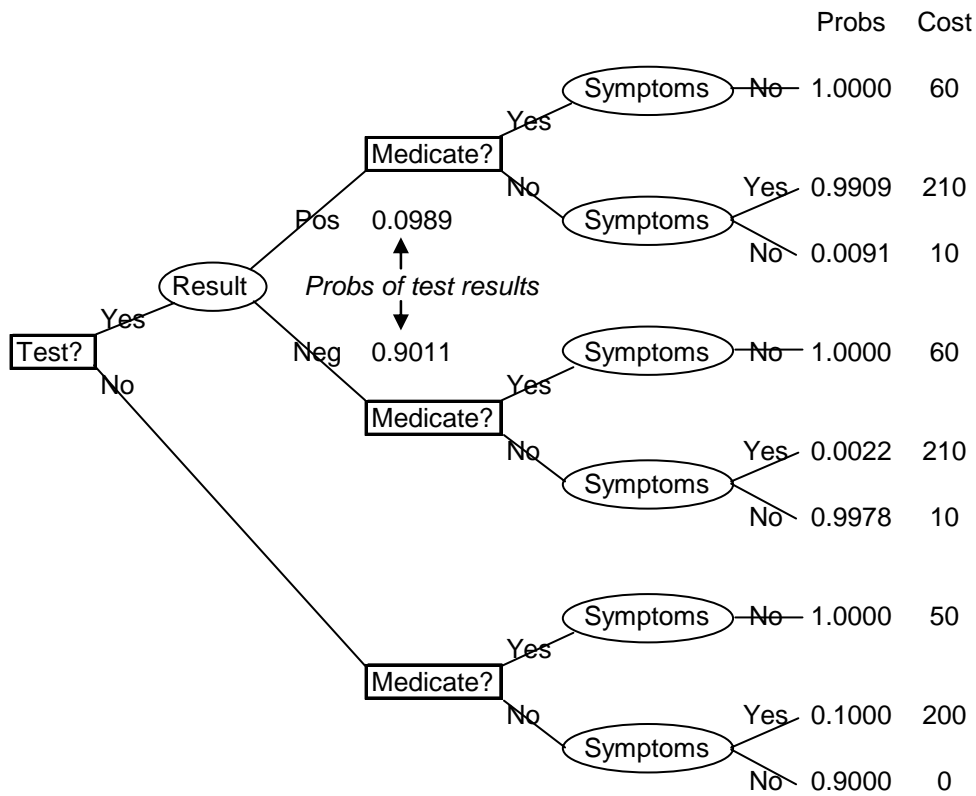
14 The diagram below is a decision problem in tree form. A patient has been exposed to a disease and may or may not be infected but is not yet manifesting symptoms. The physician can call for a diagnostic test or not. The test costs \$10 and is not perfect. Based on the test results (or not) the physician can decide to prescribe a \$50 medication that prevents the disease if it is applied before symptoms appear. If the physician waits for symptoms to appear, it is necessary to prescribe a medication that costs \$200. The prevalence of the disease is 0.15 and the test has sensitivity 0.98 and specificity 0.999. The diagram shows the decision tree in extensive form with probabilities and costs. All probability calculations have been done for you and are supplied in the diagram.

14.1 Compute the expected costs at each “Symptoms” node and write them in the first diagram on the answer sheet.

14.2 Determine the best (minimum expected cost) decision at each “Medicate?” node and write it in the second diagram on the answer sheet.

14.3 Compute the expected cost of the “Result” node and write it in the third diagram on the answer sheet.

14.4 Say if it is better (lower cost) to test or not to test.



Name: _____

Place an X in the box with the correct answer

0.0	a)	b) X	c)	d)	e)
------------	----	-------------	----	----	----

1.1	a)	b)	c)	d)	e)
------------	----	----	----	----	----

1.2	a)	b)	c)	d)	e)
------------	----	----	----	----	----

1.3	a)	b)	c)	d)	e)
------------	----	----	----	----	----

2.1	a)	b)	c)	d)	e)
------------	----	----	----	----	----

3.1	a)	b)	c)	d)	e)
------------	----	----	----	----	----

3.2	a)	b)	c)	d)	e)
------------	----	----	----	----	----

4.1	a)	b)	c)	d)	e)
------------	----	----	----	----	----

4.2	a)	b)	c)	d)	e)
------------	----	----	----	----	----

5.1	a)	b)	c)	d)	e)
------------	----	----	----	----	----

5.2	a)	b)	c)	d)	e)
------------	----	----	----	----	----

5.3	a)	b)	c)	d)	e)
------------	----	----	----	----	----

6.1	a)	b)	c)	d)	e)
------------	----	----	----	----	----

6.2	a)	b)	c)	d)	e)
------------	----	----	----	----	----

7.1	LCL:	UCL:
------------	------	------

7.2	Tail areas for t(df)		
	v = df =		
	t = z =		
	Left Tail P(t(df) ≤ z) =		
Right Tail P(t(df) > z) =			

7.3	Between:	and
------------	----------	-----

8.1	a)	b)	c)	d)	e)
------------	----	----	----	----	----

8.2	a)	b)	c)	d)	e)
------------	----	----	----	----	----

9.1	a)	b)	c)	d)	e)
------------	----	----	----	----	----

9.2	a)	b)	c)	d)	e)
------------	----	----	----	----	----

9.3	a)	b)	c)	d)	e)
------------	----	----	----	----	----

10.1	a)	b)	c)	d)	e)
-------------	----	----	----	----	----

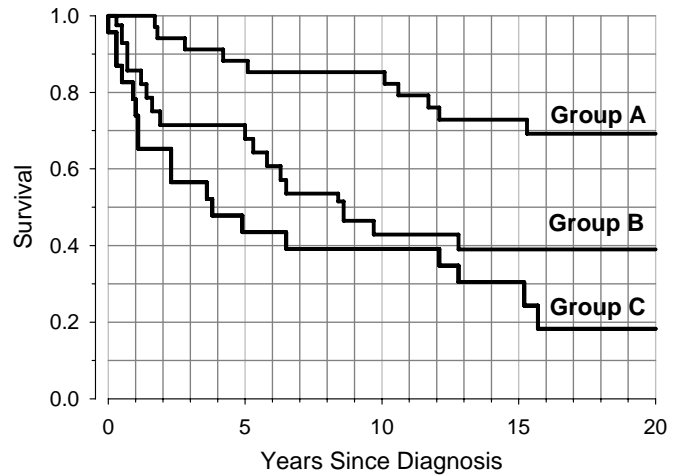
10.2	a)	b)	c)	d)	e)
-------------	----	----	----	----	----

10.3	a)	b)	c)	d)	e)
-------------	----	----	----	----	----

11.1	a)	b)	c)	d)	e)
-------------	----	----	----	----	----

11.2	a)	b)	c)	d)	e)
-------------	----	----	----	----	----

11.3	a)	b)	c)	d)	e)
-------------	----	----	----	----	----



12.1	a)	b)	c)	d)	e)
-------------	----	----	----	----	----

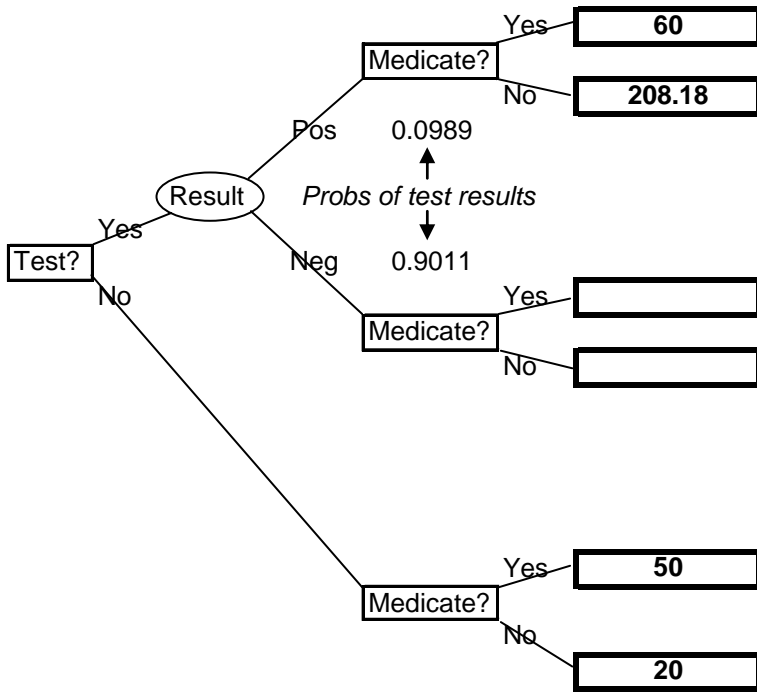
12.2	a)	b)	c)	d)	e)
-------------	----	----	----	----	----

12.3	a)	b)	c)	d)	e)
-------------	----	----	----	----	----

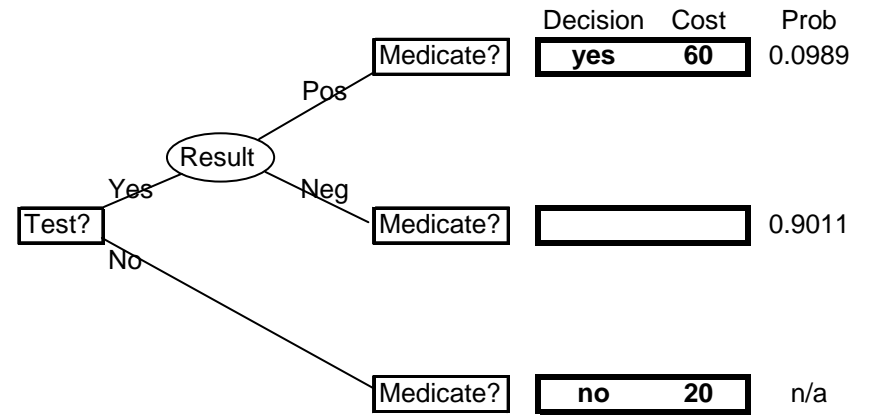
13.1	A =
	B =
	C =

(Answer 14.1, 2, 3, 4 on the back of this page)

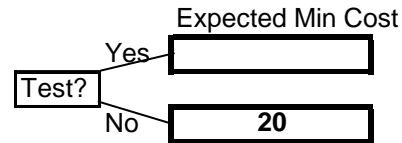
14.1 First Step: Expected Costs



14.2 Second Step: Minimum Costs & Best Decisions



14.3 Third Step: Expected Minimum Costs



14.4 Better to test?	Yes	No
-----------------------------	-----	----