Statistics 305 Final Exam

You are allowed calculators and either (i) two 8.5×11 inch formula sheets that include your statistical tables, or (ii) one 8.5×11 inch formula sheet without statistical tables. Please write your answers in an exam booklet. You may write in pen or pencil.

1. (3 marks) Short answers (no more than one sentence; 1 mark each):

   (a) What is a confounding variable?

   (b) What is an interaction variable?

   (c) A sample of 200 BC voters are asked in 2000 and again in 2010 if they would or would not vote for the BC Liberal Party. What is the appropriate test for an association between voter preference and year?

2. (5 marks) The fasting blood sugar (FBS) test is a screening test for type 1 diabetes (T1D). To evaluate the properties of the test, researchers tested 100 diabetic and 100 non-diabetic children. The results are as follows:

	T1D		
	Yes	No	Total
FBS+	86	22	108
FBS−	14	78	92
Total	100	100	200

   (a) (2 marks) Based on these data, what are the estimated sensitivity and specificity of the FBS test?

   (b) (1 mark) Why can we not estimate the positive predictive value of the FBS test directly from theses data?

   (c) (2 marks) Among Swedish children being tested in a given year, the prevalence of type 1 diabetes is about 0.0002, while in China it is about 0.00001 (note the extra zero). What are the positive predictive values of the test in these two countries?

3. (12 marks) A study of 669 children and young adults with type 1 diabetes looked at genetic risk factors for the development of insulinoma-associated (IA-2) autoantibodies, a precursor to diabetes. Subjects were classified as having high IA-2 or not. The genetic risk factor A was either present (+) or not (−). The following data were obtained:

	High IA-2	
	Yes	No
$A+$	248	111
$A-$	146	164

   (a) (1 mark) What are the proportions of subjects with high IA-2 in the $A+$ and $A-$ groups, respectively?

(b) (2 marks) Give a 95% confidence interval for the difference in population proportions of high IA-2 in the $A+$ and $A-$ groups.

(c) (2 marks) Report an approximate p-value for the test of the hypothesis that the difference between these two population proportions is zero versus the two-sided alternative. Your calculations should show the SE used in constructing the test statistic.

(d) (6 marks) Now use the chi-square test to test for an association between high IA-2 status and genetic risk factor status at the 5% level. Your calculations should show the table of expected cell counts and the table of chi-square contributions.

(e) (1 mark) Write a sentence to report the conclusion of your chi-square test.

4. (18 marks) This question uses the low birth weight data. To investigate the difference in systolic blood pressure (sbp) by gender (sex), you are given the following summary statistics:

```
sex=0 (female) n = 56, mean = 46.46429, SD = 11.14526
sex=1 (male) n = 44, mean = 47.86364, SD = 11.8057
```

For the two-sample t-procedures that do not assume equal SD, computer software estimates the df to be 89.86.

(a) (4 marks) Construct a 95% CI for the difference in sbp by sex.

(b) (2 marks) Using Table A-4 compute an approximate p-value for the test of the null hypothesis of no difference in sbp by gender against the two-sided alternative.

(c) (2 marks) Are our rules of thumb for safe use of the t-procedures met? If so, why? If not, why not?

(d) (2 marks) We will now investigate the association between sbp and five minute apgar score (apgar5), accounting for sex as necessary. The first step is to fit a multiple regression model that includes interaction between sex and apgar5. The fitted model summary is:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  42.9994     4.0178  10.702   <2e-16
apgar5        0.5641     0.6050   0.932    0.353
sex          -0.7292     6.5296  -0.112    0.911
apgar5:sex    0.3118     0.9710   0.321    0.749
```

Test the hypothesis of no interaction versus the hypothesis of interaction at the 5% level and write a sentence to report your conclusions.

(e) (2 marks) Based on the following two fitted models, would we consider sex to be a confounder of the relationship between apgar5 and sbp? If so, why? If not, why not? (**NB1:** I'm asking if sex is a confounder, not whether apgar5 is a confounder as in the practice problems. **NB2:** The p-values for the tests of each coefficient have been intentionally omitted.)

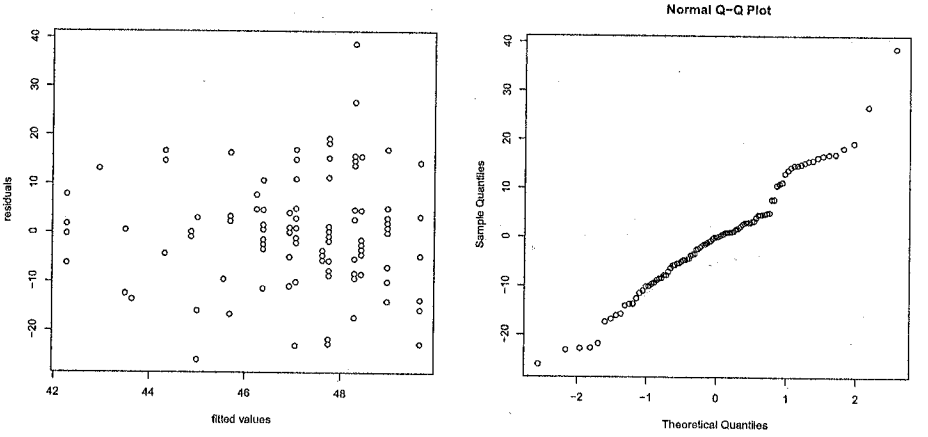
```
Coefficients:
            Estimate Std. Error t value
(Intercept)  42.2559     3.2683  12.929
apgar5        0.6851     0.4710   1.455
sex           1.2325     2.2945   0.537

Coefficients:
            Estimate Std. Error t value
(Intercept)  42.7192     3.1410  13.601
apgar5        0.6977     0.4687   1.489
```

(f) (3 marks) In the model with apgar5 as the only explanatory variable, report the estimated apgar5 effect, a 95% CI and a test of the hypothesis of no sex effect against the two sided alternative at the 5% level.

(g) (1 mark) Interpret the estimated apgar5 effect from the model with only apgar5.

(h) (2 marks) Examine the following residual diagnostics and comment on the three assumptions for inference. You must name each assumption to get full marks.



5. (10 marks) In the type 1 diabetes data of problem 3, age of the subject at the time of diagnosis was also available. Define the following variables:

   - **ia2** is 1 if the subject has high IA-2 autoantibodies and 0 otherwise,
   - **genet** is 1 if the subject has the genetic risk factor and 0 otherwise, and
   - **age** is the age of the subject in years.

   The association of primary interest is between ia2 and genet, but we will consider the impact of age as necessary.

(a) (1 mark) A summary of the fitted model that includes interaction between genet and age is as follows:

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.14976    0.54790   5.749 8.99e-09
genet       -12.17937    1.23491  -9.863  < 2e-16
age          -0.23979    0.03919  -6.118 9.48e-10
genet:age     0.75824    0.71380   1.062    0.288
```

Test the hypothesis of no interaction versus the hypothesis of interaction at the 5% level and write a sentence to report your conclusions.

(b) (2 marks) Based on the following two fitted models, would we consider age to be a confounder of the relationship between genet and ia2? If so, why? If not, why not?

```
Coefficients:
            Estimate Std. Error z value
(Intercept) -1.01864    0.32768  -3.109
genet        0.51681    0.20932   2.469
age          0.06593    0.02238   2.946
```

```
Coefficients:
            Estimate Std. Error z value
(Intercept)  -0.1163     0.1138  -1.022
genet         0.9202     0.1612   5.708
```

(c) (2 marks) In the model with genet and age (but not their interaction) as explanatory variables, report and interpret the estimated genetic effect in terms of the odds of high IA-2 status.

(d) (2 marks) In the model with genet and age (but not their interaction) as explanatory variables, report and interpret the estimated age effect in terms of the odds of high IA-2 status.

(e) (2 marks) Report a 95% CI for the genetic OR.

(f) (1 mark) Test the hypothesis of no genetic effect against the two sided alternative at the 5% level.