The final examination for STAT1201 in Semester 1, 2011 will be very different to previous exams in the course. This sample exam aims to give you a guide to the scope, style and structure of the final exam.

**Question 1**

Alice has seven close friends, three male and four female, who are all about the same height as each other. She is wondering if the males around this height tend to be heavier than females on average. She sets up a scale in her floor and secretly weighs each friend when they visit. The resulting weights are

<table>
<thead>
<tr>
<th>Male</th>
<th>Female</th>
<th>Female</th>
<th>Male</th>
<th>Female</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>56</td>
<td>39</td>
<td>68</td>
<td>50</td>
<td>86</td>
<td>70</td>
</tr>
</tbody>
</table>

a) State in words the null hypothesis ($H_0$) and alternative hypothesis ($H_1$) that Alice is interested in testing. [1 mark]

$H_0$: 

$H_1$: 

b) What is the observed difference in mean weight between males and females? [1 mark]

c) Calculate the exact $P$-value for a randomization test of $H_0$ and interpret the result. Show your working. [4 marks]

**Question 2**

Suppose a couple decide to have children until they either have a child of each sex or they have three children.

a) Using $G$ for girl and $B$ for boy, write down the sample space of possible outcomes that can arise from this process. For example, $GGB$ could be the outcome where the couple has two girls and then a boy. [1 mark]

b) Define a probability function for this process. State any assumptions you make about the process. [2 marks]

c) Now let the random variable $X$ be the number of girls resulting from this approach to having children. Write down the sample space for $X$ and give the probability function for $X$ based on your answer to (b). [1 mark]

d) What is the expected number of girls, $E(X)$? [1 mark]

e) What is the standard deviation of the number of girls, $sd(X)$? [1 mark]
Question 3

a) The estimated odds of finding a question on the exam about odds are 2.6 to 1. What is the corresponding probability? [1 mark]

b) Suppose $X$, the age of a randomly chosen STAT1201 student, has the following probability density curve:

![Probability Density Curve]

What is the probability that a randomly chosen student is less than 20 years old? [1 mark]

Question 4

For a particular population group it is estimated that the prevalence of prostate cancer is 15.2%. A new diagnostic procedure for this disease has a sensitivity of 0.951, correctly diagnosing an individual with prostate cancer 95.1% of the time. The specificity of the procedure is 0.976, so that an individual without the disease is correctly diagnosed 97.6% of the time.

Suppose the diagnostic procedure indicates an individual from this population group has the disease. What is the probability that they actually do have prostate cancer? [2 marks]

Question 5

It has come to light in recent years that drug companies have paid groups of medical scientists as “ghostwriters” to write opinion pieces about the current role of particular drugs in the treatment of particular diseases. At the same time a leading doctor in the disease area is approached and asked whether they would like to write an opinion piece on the drug with research assistance from a group of medical scientists. If the doctor agrees they are then sent the draft with the freedom to edit the draft however they like and then submit it to a leading journal. This is often done by the doctor without acknowledgment of the “assistance” and background - the sole author listed is the doctor.

Explain briefly what, if anything, is ethically wrong with this practice. If you do not think there is anything ethically wrong with this practice, provide a brief defense of it. [4 marks]
Question 6

a) Two numbers have sample mean 45.6 and sample standard deviation 5.09. What are the two numbers? [2 marks]

b) The weights of bananas in a large crate have a Normal distribution with \( \mu = 163 \) grams and \( \sigma = 18.7 \) grams. Consider the random process of picking three bananas and putting them in a bag. What is the standard deviation of the total weight of the three bananas? [1 mark]

Question 7

Researchers conducted an experiment to see whether drinking beer increases pulse rate. A sample of 14 subjects had their resting pulse rate recorded. They were each then given 250 mL of beer to drink. After remaining seated for 30 minutes, their pulse rate was measured again.

The researchers observed a mean increase in pulse rate of 3.1 beats per minute (bpm) with a standard deviation of 5.91 bpm.

a) State the null and alternative hypotheses of interest to the researchers. [1 mark]

\[ H_0: \]

\[ H_1: \]

b) Calculate the outcome of a \( t \)-test using the given values. What do you conclude? [2 marks]

Question 8

Researchers are interested in factors involved with low birth weight deliveries at an obstetrics clinic. Based on a sample 95 births at the clinic, the following table gives data on whether the birth was low weight (under 2500 g) and whether the mother had a history of hypertension.

<table>
<thead>
<tr>
<th></th>
<th>Low birth weight</th>
<th>Normal birth weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>No hypertension</td>
<td>31</td>
<td>38</td>
</tr>
</tbody>
</table>

Is there any evidence of an association between low birth weight deliveries and maternal hypertension? Show your working. [4 marks]
Question 9

Researchers collected 20 juvenile hammerhead sharks from turbid waters to study how their ocular tissues change when exposed to solar radiation. Half of the sharks were kept in a shallow, outdoor pen, with high light levels, while the other half were kept in a similar pen but with constant shade. Immediately after capture and then on days 7, 14, 20, and 27, two sharks from each group were removed from their holding pens and euthanized. After dissecting the left eye from each euthanized shark, light transmission through the cornea was measured at various wavelengths. Along with the days of exposure, the researchers recorded the wavelengths (nm) that gave the maximum transmission for the cornea for each shark, as well as the sex of the shark (male/female) and its fork length (cm). The following table shows the data for the 10 sharks exposed to sunshine.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Female</th>
<th>Female</th>
<th>Female</th>
<th>Female</th>
<th>Male</th>
<th>Male</th>
<th>Male</th>
<th>Male</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>14</td>
<td>14</td>
<td>20</td>
<td>20</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Fork Length</td>
<td>47.0</td>
<td>45.1</td>
<td>45.2</td>
<td>48.4</td>
<td>42.9</td>
<td>40.4</td>
<td>39.8</td>
<td>40.6</td>
<td>44.9</td>
<td>42.6</td>
</tr>
<tr>
<td>Cornea</td>
<td>345.8</td>
<td>337.8</td>
<td>342.3</td>
<td>344.9</td>
<td>348.6</td>
<td>357.3</td>
<td>357.7</td>
<td>350.2</td>
<td>346.9</td>
<td>348.9</td>
</tr>
</tbody>
</table>

The following figures show a scatter plot of the relationship between maximum cornea transmission wavelength and days of exposure to light along with a Normal quantile plot of the residuals from the least-squares line.

![Scatter plot](image1)

a) Enter the appropriate \( t \) values in the two blank boxes in the following regression analysis [1 mark]

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>( t ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>343.4356</td>
<td>3.0025</td>
<td></td>
</tr>
<tr>
<td>Days</td>
<td>0.3386</td>
<td>0.1811</td>
<td></td>
</tr>
</tbody>
</table>

b) Based on the table in (a), give bounds on the \( P \)-value for the test of linear association between maximum cornea transmission wavelength and days of exposure to light. What do you conclude? Comment on the assumptions underlying your inference. [3 marks]

c) The residual standard error for this analysis was 5.429 nm. The mean number of days was 13.6 with standard deviation 9.99. Give a 95% confidence interval for the mean cornea transmission for sharks exposed to 30 days of sunshine. [3 marks]
Question 10
A genetically engineered (GE) tomato has been shown to be resistant to leaf curl virus but its safety for human consumption needs to be assessed. One experiment for the assessment studied 30 mice over six weeks. Wild-type and GE tomato were freeze dried and ground to a fine powder. Ten of the mice had 100 mg/kg body weight of wild-type powder added to their regular diet, ten had 100 mg/kg body weight of the GE powder, while the remaining ten were used as a control. Initial body weights (g) were recorded along with the final body weight (g), total plasma protein (g/dL), total and HDL cholesterol levels (mg/dL) and serum glutamic oxaloacetic transaminase (SGOT; IU/L) at the end of the six weeks.

a) With three groups of 10 subjects each, the power of an F-test to detect a difference in means with an $R^2$ value of 0.269 was found to be 0.8096. In this case what is the probability of making a Type II error? [1 mark]

b) Complete the seven blank boxes in the following ANOVA table for testing for a difference in mean HDL cholesterol between the three diets. [2 marks]

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residuals</td>
<td>1165.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1258.34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

c) What is the $R^2$ value for this model of mean HDL cholesterol? [1 mark]

d) Give bounds on the $P$-value for the $F$ statistic in the ANOVA table. What do you conclude? [1 mark]
Question 11

A genetic study into psychiatric disorders investigated a single nucleotide polymorphism (SNP) of the DTNBP1 gene. The subjects were 45 schizophrenia patients and 52 unrelated Caucasian controls. Researchers recorded the sex and age of the subjects and then determined their SNP genotype from a saliva sample as AA, AC or CC. To help establish how the polymorphism might be related to psychiatric disorders, the researchers also made a measurement of the glutamine/glutamate ratio in the parieto-occipital cortex using magnetic resonance spectroscopy.

a) In a case-control study, the two groups should be homogenous with respect to variables other than the predictor of interest. Here the mean age of subjects in the control group was 31.2 years with a standard deviation of 6.72 years. In the patient group the mean age was 29.8 years with a standard deviation of 5.9 years. Is this any evidence of a difference in mean age between the groups? Show your working and P-value bounds. [3 marks]

b) The researchers wanted to know if the glutamine/glutamate ratio in the parieto-occipital cortex could be used as a diagnostic predictor for schizophrenia. A logistic regression in R for the relationship between being in the schizophrenia group and the glutamine/glutamate ratio gave the following results:

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.4719</td>
</tr>
<tr>
<td>Ratio</td>
<td>4.3771</td>
</tr>
</tbody>
</table>

Is there any evidence of an association between being in the schizophrenia group and the glutamine/glutamate ratio? Show your working and P-value. [2 mark]

c) Interpret the coefficient of Ratio in the logistic regression in (b). [1 mark]

d) Based on the logistic regression in (b), what is the estimated probability of being in the schizophrenia group for a glutamine/glutamate ratio of 0.63? [1 mark]

e) Based on the logistic regression in (b), what is the glutamine/glutamate ratio at which an individual would have an estimated 50:50 chance of being in the schizophrenia group? [1 mark]