Stat 401 **FINAL EXAM** **May 18, 2006**

J.Millson

1. In a study designed to investigate the effects of a strong magnetic field on the early development of mice, 5 cages, each containing a young mouse, were subjected for a period of 5 days to a magnetic field. Five other young mice, housed in 5 similar cages, were not put in the magnetic field and served as controls. The total weight gains for each mouse for the first set were 23 milligrams, 19 milligrams, 24 milligrams, 27 milligrams and 21 milligrams. The total weight gains for the second set were 23.5 milligrams, 19.5 milligrams, 23.8 milligrams, 26.5 milligrams and 22 milligrams.

(a) Test whether the variances of the weight gains in the two families of mice are equal using the two-sample F-test on your calculator. Make your decision on the basis of whether or not the resulting P-value for the F-test is large or small.

(b) Do a two-sample t-test to test whether exposure to a magnetic field influences the weight gain of a young mouse (use the level $\alpha = .05$). Use your answer to (a) to decide whether you should do a pooled or unpooled two-sample t-test.

(20 points)

2. A car manufacturer is trying to decide which of three catalytic converters to install on its new models. All three meet the government emission standards but they may not all give the same mileage. Twelve new cars, all alike, are selected for the test. Converters A, B and C are each installed on four of the cars. The cars are then driven over a specially engineered track that simulates city driving. The mileage estimates are listed below.

A	21.6	23.2	20.5	21.7
B	22.8	25.6	24.7	25.0
C	23.9	24.6	23.8	24.0

Test whether all three converters are equally economical at level $\alpha = .05$.

(10 points)

3. The soils of 4 lots of seedlings were planted with varying mixtures of nutrient I and nutrient II. After three weeks the seedlings were measured and their average heights recorded. The data is listed below.

Plot	Height	Nutrient I	Nutrient II
1	2.4	1	1
2	3.5	2	1
3	4.6	3	1
4	3.0	1	2

Assuming that the average height z of seedlings is linearly dependent on the amount x of nutrient I and the amount y of nutrient II, find the least squares estimate of this average height.

(This is *multiple* regression. Multiple regression is not available on your calculator. The formula is given in (13.18) from the text, pg. 578 in Edition 5 and pg. 582 in Edition 6. Since the general formula from the text is hard to understand I will give you the formula you need for this case below. In this case you are looking for the least squares *plane* $z = c + ax + by$ instead of the least squares *line* $y = b + ax$.)

Theorem 1. *The least square estimators* $\hat{a}$, $\hat{b}$ *and* $\hat{c}$ *for* a, b *and* c *are the solutions of the following three-by-three linear system*

$$\begin{aligned} 4\ \hat{c} + \sum_{i=1}^{4} x_i\ \hat{a} + \sum_{i=1}^{4} y_i\ \hat{b} &= \sum_{i=1}^{4} z_i \\ \sum_{i=1}^{4} x_i\ \hat{c} + \sum_{i=1}^{4} x_i^2\ \hat{a} + \sum_{i=1}^{4} x_i y_i\ \hat{b} &= \sum_{i=1}^{4} x_i z_i \\ \sum_{i=1}^{4} y_i\ \hat{c} + \sum_{i=1}^{4} x_i y_i\ \hat{a} + \sum_{i=1}^{4} y_i^2\ \hat{b} &= \sum_{i=1}^{4} y_i z_i \end{aligned}$$

First find the coefficients $\sum_{i=1}^{4} x_i$, $\sum_{i=1}^{4} y_i$, $\sum_{i=1}^{4} z_i$, $\sum_{i=1}^{4} x_i^2$, $\sum_{i=1}^{4} y_i^2$, $\sum_{i=1}^{4} x_i y_i$, $\sum_{i=1}^{4} x_i z_i$ and $\sum_{i=1}^{4} y_i z_i$ of the above equations for the unknowns $\hat{a}, \hat{b}, \hat{c}$. Then solve the resulting linear system. Your calculator will solve the equations as I now explain or you can solve them by hand. You will get 20 points for writing down the correct linear system and 10 points for solving the equations.

To solve a system of linear equations $Ax = b$ with A a 3 by 3 matrix (so b is an 3 by 1 matrix) using your calculator procede as follows. First enter the matrix A by going to MATRIX (the third key down on the left-hand side). Then EDIT $\rightarrow$ [A] $\rightarrow 3 \times 3$ (you have to change the 1×1 to 3×3). Now enter the entries of A. Enter the right-hand side b of the equation as a 3×1 matrix B in the same way . Now to solve the equation you have to compute $A^{-1} \times B$. To do this MATRIX $\rightarrow$ [A] $\rightarrow A^{-1}$ (you get this by hitting the key x^{-1}) $\rightarrow A^{-1}\times$ (you get this by hitting the $\times =$"times" key) $\rightarrow A^{-1} \times B$. You get this by pasting in B via MATRIX $\rightarrow$ [B].

(30 points)

4. The point of the following problem is to find a confidence interval and a test for β_0, the y- *intercept* of the regression line. The theory will be very similar to the one we developed in class for β_1, the *slope* of the regression line.

Let $Y \sim N(\beta_0 + \beta_1 x, \sigma^2)$. Suppose that $(x_1, Y_1), (x_2, Y_2), \cdots, (x_n, Y_n)$ is a random sample from the space of Y (i.e. we are in the framework of simple linear regression). Define a random variable $S_{\widehat{\beta_0}}$ by the formula

$$S^2_{\widehat{\beta_0}} = \left(\frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \overline{x})^2}\right) \quad S^2$$

Here we have

$$S^2 = \frac{1}{n-2} SSE = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2.$$

Then we define $S_{\widehat{\beta_0}}$ to be the square root of $S^2_{\widehat{\beta_0}}$. Assume the following theorem from probability theory.

Theorem 2. *The random variable* $T = \frac{\widehat{\beta_0} - \beta_0}{S_{\widehat{\beta_0}}}$ *has* t- *distribution with* $n-2$ *degrees of freedom.*

(a) Write down a two-sided $100(1-\alpha)\%$ confidence interval for β_0.

(b) Prove your interval has the correct confidence level.

(c) Write down a decision rule for testing

$$H_0 : \beta_0 = (\beta_0)_0$$
against
$$H_a : \beta_0 \neq (\beta_0)_0.$$

(d) Prove that your test has significance level α.

(20 points)

5. In this problem you will derive the formula for β for the upper-tailed one sample t-test.

We wiil assume the following theorem.

Theorem 3. *Suppose that Y has normal distribution with mean Δ and variance 1 and V has chi-squared distribution with d degrees of freedom and Y and V are independent. Then the ratio $T = Y/(\sqrt{V/d})$ has noncentral t-distribution with noncentrality parameter Δ and d degrees of freedom.*

We will abbreviate this to

$$T \sim t_{\Delta,d}.$$

We will let $F_{T_{\Delta,d}}(t)$ denote the cumulative distribution of a random variable with the above distribution.
Suppose that $X_1, X_2, \cdots, X_n$ is a random sample from a normal distribution with mean μ and variance σ^2. We want to do a t-test for

$$H_0 : \mu = \mu_0$$
against
$$H_a : \mu > \mu_0.$$

We use the decision rule: reject H_0 if

$$\overline{x} \geq \mu_0 + t_{\alpha,n-1}\frac{s}{\sqrt{n}}.$$

(a) Suppose that in fact we have $\mu = \mu' > \mu_0$ (so H_a is true). Show that the ratio $T = (\overline{X} - \mu_0)/(S/\sqrt{n})$ has noncentral t-distribution with noncentrality parameter $\Delta = \frac{\mu' - \mu_0}{\sigma/\sqrt{n}}$ and $n-1$ degrees of freedom.

(Hint: this is an exercise in fractions similar to what we did in class to get the ordinary t-distribution. You will need $E(\overline{X} - \mu_0) = \mu' - \mu_0$ and $V(\overline{X} - \mu_0) = V(\overline{X}) = \frac{\sigma^2}{n}$.)

(b) Assume $\mu = \mu' > \mu_0$. Prove that the probablity $\beta(\mu')$ of (incorrectly) accepting H_0 using the above decision rule is given by

$$\beta(\mu') = F_{T_{\Delta,n-1}}(t_{\alpha,n-1})$$

where Δ is as in (a).

You can still do (b) even if you didn't get (a). Just treat (a) as a known theorem, then (b) is a standard computation of an error probability.

(20 points)