# Applied Statistics

# MS Comprehensive Examination

## April 23, 2011

***Instructions:***

Please answer all questions. Points are as noted.

Record your answers in your blue books.

You may use your books, notes and calculator for this exam.

You have three hours.

1. (10 points) Birthweight Problem: The data are the birthweights ($y$, in grams) and estimated gestational age ($x$, in weeks) for 12 female babies born in a certain hospital. You may regard these 12 babies as a random sample of all female babies born at the hospital. Consider the model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i .$$

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 1 | 616401 | 616401 | 24.78 | 0.0006 |
| Error | 10 | 248726 | 24873 | | |
| Corrected Total | 11 | 865127 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 157.71045 | R-Square | 0.7125 |
| Dependent Mean | 2911.33333 | Adj R-Sq | 0.6837 |
| Coeff Var | 5.41712 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | 95% Confidence Limits | |
|---|---|---|---|---|---|---|---|
| Intercept | 1 | -2141.66667 | 1016.04886 | -2.11 | 0.0613 | -4405.56461 | 122.23128 |
| age | 1 | 130.40000 | 26.19428 | 4.98 | 0.0006 | 72.03551 | 188.76449 |

Use the SAS output given above to answer following questions.

a. Find $\hat{\beta}_0$ and $\hat{\beta}_1$.

b. Test $H_0 : \beta_1 = 0$ vs $H_\alpha : \beta_1 \neq 0$.

c. Find the 95% confidence interval for $\beta_1$.

d. Find SSR, SSE and SST.

e. Find $R^2$.

2. (20 points) The following table provides data on body weight gain (g) for two independent samples: a sample of animals given a 1 mg/pellet dose of a certain soft steroid and a sample of control animals:

| | | | | | |
|---|---|---|---|---|---|
| Steroid | 33 | 45 | 42 | 41 | 39 |
| Control | 40 | 31 | 38 | 34 | 35 |

Use Wilcoxon rank sum test at level $\alpha = 0.05$ to decide whether there is a significant difference in weight gain between steroid and control groups. Find the exact p-value and use it to make your decision.

3. (20 points) In a study of a sample of people on whether they report drinking alcohol frequently (1 = yes, 0 = no). Fit logistic regression model with Extroversion/Introversion (E/I) and Thinking/Feeling (T/F) as predictors.

Analysis Of Parameter Estimates

| Parameter | | DF | Estimate | Standard Error | Likelihood Ratio 95% Conf. Limits | | Wald Chi-Square |
|---|---|---|---|---|---|---|---|
| Intercept | | 1 | −2.8291 | 0.1955 | −3.2291 | −2.4614 | 209.37 |
| EI | e | 1 | 0.5805 | 0.2160 | 0.1589 | 1.0080 | 7.22 |
| TF | t | 1 | 0.5971 | 0.2152 | 0.1745 | 1.0205 | 7.69 |

LR Statistics

| Source | DF | Chi-Square | Pr > ChiSq |
|---|---|---|---|
| EI | 1 | 7.28 | 0.0070 |
| TF | 1 | 7.64 | 0.0057 |

Use the SAS output given above to answer following questions.

a. Find the estimate of the probability of drinking alcohol frequently for someone of personality type introverted and feeling.

b. Report and interpret the estimated conditional odds ratio between E/I and the response.

c. Use the limits reported to construct a 95% likelihood-ratio confidence interval for the conditional odds ratio between E/I and the response.

d. Test whether E/I has an effect on the response, controlling for T/F.

4. (15 points) Following are some statistics from data from a normal population. In this problem, we will use this data to test $H_0$: $\sigma^2 = 16$ versus $H_A$: $\sigma^2 > 16$ with level of significance $\alpha = .05$.

| Variable | N | Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
|---|---|---|---|---|---|---|---|---|
| Data | 30 | 6.227 | 5.167 | -3.002 | 3.256 | 5.059 | 8.695 | 22.666 |

Do this two ways:

a. the usual way using the exact chi-square distribution. Give your test statistic, the critical value, your decision, and bounds on the P-value provided by the attached chi-square table.

b. using the appropriate normal approximation to the chi-square. Give your test statistic, the critical value, your decision, and the P-value provided by the attached normal table. To use the normal approximation, recall these facts regarding the chi-square distribution:

i) If W ~ $\chi^2_\nu$ then E(W) = $\nu$ and Var(W) = $2\nu$ and

ii) If $\nu$ is large, then W is approximately normal.

c. Comment on the similarities and differences between your answers in parts a and b. In particular, discuss whether or not the normal approximation is justified and the impact of this on your answer to part b.

5. (15 points) We want to do a paired difference test of the equality of means for a bivariate normal population where the two variables have equal variances and correlation about .75. Say that a pilot study has produced data with the following statistics for the individual variables:

| Variable | N | Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
|---|---|---|---|---|---|---|---|---|
| X | 10 | 4.105 | 1.157 | 2.833 | 2.998 | 3.766 | 5.097 | 5.983 |
| Y | 10 | 4.590 | 1.219 | 2.853 | 3.653 | 4.501 | 5.628 | 6.526 |
| X-Y | 10 | -0.485 | 0.959 | -2.051 | -1.572 | -0.161 | 0.213 | 0.776 |

Correlation between X & Y = 0.676.

a. First use this pilot data to test the null hypothesis that the means of X & Y are equal versus the one-tailed alternative that Y has the greater mean. Use level of significance $\alpha = .05$.

b. Further say that a scientifically meaningful difference between the two means is $\mu_Y - \mu_X = 0.6$. Find the sample size required to detect this difference with a power of 0.90. Use $\alpha = 0.05$. Use the statistics given as a pilot study where necessary. Clearly state your assumptions.

c. Comment on the large sample method you used to derive n in part b relative to the n that you found. Is your n large enough to justify having used the normal approximation? Use your sample size from part b to find the actual critical value. Then what can you say about the true power at 0.6?

6. (20 points) In this problem we will explore the question of sample size for one-tailed tests regarding the variance of a normal population. Assume that $X_1,\ldots,X_n \sim$ i.i.d. $N(\mu,\sigma^2)$ and that we want to test $H_0$: $\sigma^2 = \sigma_0^2$ versus $H_A$: $\sigma^2 > \sigma_0^2$ with level of significance $\alpha$. Further say that we want to find the rejection region and sample size so that the power of the test at a specific variance $\sigma_A^2 > \sigma_0^2$ is equal to $1-\beta$. To do this, first recall these facts regarding the chi-square distribution:
   i) If $W \sim \chi_\nu^2$ then $E(W) = \nu$ and $Var(W) = 2\nu$ and
   ii) If $\nu$ is large, then W is approximately normal.

   a. Write down the usual test statistic for doing this test and identify its distribution under $H_0$. For the sake of future discussion, let's name this statistic W.

   b. Call the critical value for this test "c". Give the form of the rejection region in terms of W and c.

   c. Let $z_\alpha$ denote the value such that $P(Z > z_\alpha) = \alpha$, where $Z \sim N(0,1)$. Assume that n will be large enough and use the normal approximation to relate c to $z_\alpha$.

   d. Write down the exact probability equation for $\beta$ in terms of W, $\sigma_A^2, \sigma_0^2$ and c.

   e. Again assume that n is large enough and use your answer in part d to relate $\sigma_A^2, \sigma_0^2$, n, c and $z_\beta$.

   f. In parts c and e you have two equations in the two unknowns, c and n. Use these to show that c and n depend upon $\sigma_A^2$ & $\sigma_0^2$ only through their ratio $r = \sigma_A^2 / \sigma_0^2$.

   g. To solve these equations, it might be helpful to use r from part f and $q = z_\alpha + z_\beta$. I recommend: solve for c in terms of n first. Do this. So you can check your answer, I get $c = \dfrac{q\sqrt{2(n-1)}}{1-r}$. If you can't get this answer, use this formula for parts h, I & j anyway.

   h. Now solve for c and n. Hint: Use the equations from parts c and g.

   i. If $\alpha = .05$ and $\beta = .10$ (power = .9) and $\sigma_A^2 = 4\sigma_0^2$ (r=1/4, or the ratio of standard deviations is 2), find n and c.

   j. Check this answer using the chi-square table. Comment on any inaccuracy based on the fact that you derived these values using a normal approximation.