AMS 102.4 – Midterm 1 **FORM A** **NAME:_____**
10/12/2011 **ID #: _____**

1) (16 points) Determine whether each statement is true or false, if a statement is false, provide an explanation as to why it is false.

a) A Type II error can only occur when the alternative hypothesis is the underlying truth.

b) Response bias is where a certain type of unit is chosen for the sample that will provide data that are in favor of the study being run, hence they respond as the surveyor wants.

c) A retrospective study usually gives more accurate results than a prospective observational study since the events have already happened.

d) When combining individual strata's estimates into an overall estimate, we use a weighted average with the weights equal to the sample size we took inside each strata divided by the overall sample size.

For Example: $26 = \frac{10}{50}(25) + \frac{15}{50}(15) + \frac{15}{50}(35) + \frac{10}{50}(30)$

Strata	*Strata Size*	*Sample Size*	*Statistic*
Strata 1	100	10	25
Strata 2	300	15	15
Strata 3	400	15	35
Strata 4	200	10	30
TOTAL	**1000**	**50**	**26**

e) Histograms are a good method for displaying qualitative data such as the numbering of rooms in a school building.

f) Observational Studies are the most effective way to establish causation between an explanatory variable and response variable

2) (20 points) You are asked to study the relationship between years of experience and salary for employees at a small firm. The data is presented in the table below

Position	*Yrs Experience*	*Salary*	*Position*	*Yrs Experience*	*Salary*
Analyst	5	$42,500	Analyst	3	$41,500
Analyst	3	$40,000	Associate	4	$52,000
Analyst	2	$42,500	Associate	12	$57,500
Analyst	5	$46,000	Manager	10	$70,000

a) (8 points) Calculate the mean and variance for Yrs of Experience.

b) (8 points) Calculate the 5-number summary for salary, draw a modified boxplot (using the 1.5 IQR rule)

c) (4 points) Make a scatterplot of the relationship under study and comment on the type of relationship (form, direction, strength)

3) (24 points) You currently are unhappy with your commute to work. On average it takes you 30 minutes and you want to try to find new routes that will take you less time. You've only been working at your job for 4 weeks and the driving times to work are:

25 31 22 29 29 29 33 31 37 30

35 29 30 30 30 28 36 26 28 32

So tomorrow you plan on using a different route and you want to determine statistically if you should permanently change your route.

a) (6 points) What are your null and alternative hypotheses?

b) (4 points) Tomorrow you drive to work and it takes you 28 minutes. What is the p-value associated with this value?

c) (4 points) After reviewing your results, you realized you forgot to stop the clock where you usually do and it really only took you 27 minutes to get to work. What is the p-value of this new value?

d) (7 points) If the probability of a type I error was set to 10%, can you make a decision make based on the sampled value of 27 minutes, or do you need more information? If you can make a decision, what is it? If you need more information, explain what information you need?

e) (3 points) Provide at least two potential confounding variables that could influence the accuracy of your results?

*******Extra Credit (5 pts)**

We discussed in class that a linear transformation is defined as taking each data point x_i in our sample and converting it into another data point y_i by using the transformation $y_i = mx_i + b$ for a given m and b (i.e. if we want to convert Celsius data points into Fahrenheit data points we use the formula $y_i = \frac{9}{5}x_i + 32$).

1) (2 points) If the sample mean of the non-transformed observations is given as $\bar{x}$, then the transformed sample mean is given by $\bar{y} = m\bar{x} + b$. Prove this relationship is true for any m and b. (Hint: substitute $y_i = mx_i + b$ into the average formula for y).

2) (3 points) If the sample standard deviation of the non-transformed observations is given as s_x then the transformed sample standard deviation is given by $s_y = |b|s_x$. Prove this relationship is true for any m and b. (Hint: use the original sample variance formula $s_y = \frac{1}{n-1}\sum(y_i - \bar{y})^2$ and substitution)