Name: Student Number:

STAT 285:

Final Examination

Instructions: This is an open book test. As such you may use formulas such as those for means and variances of standard distributions from the book without deriving them. You may use notes, text, other books and a calculator; you may not use a computer. Your work will be marked for clarity of explanation. I expect you to explain what assumptions you are making and to comment if those assumptions seem unreasonable. In general you need not finish doing arithmetic in confidence interval problems; I will be satisfied if your answers contain things like

$$27 \pm 1.96\sqrt{247.5/11},$$

but I have to be absolutely convinced you know what arithmetic to do! In hypothesis testing problems you will have to finish the arithmetic enough to reach a real world conclusion. I want the answers written on the paper. The exam is out of **70**.

1. We generally assume that when a coin is tossed it lands heads up with probability 0.5. It has been suggested, however, that if you do the experiment in a different way the chance might change. One such way is to stand the coin on edge on a hard flat surface, hold it upright with a finger and then flick the edge with a finger to send the coin spinning away.

   A group of 107 statistics students actually did this, 40 times each for a total of 4280 spins.

   (a) [**5 marks**] Suppose they got a total of 2376 heads. Give a 99% confidence interval for the probability that spinning produces heads.

(b) **[5 marks]** Is it reasonable to believe that this method produces heads with probability 1/2?

2. **[5 marks]** It has also been suggested that if you flip a coin in the usual way and catch it in your hand it is slightly more likely than 50% to land the same way up as it started. If the chance of landing the same side up as it started were really 0.51 how many tosses would I need to make to have at least a 90% chance that a level 5% test would detect a significant difference between the probability of heads and 0.5?

3. A large population of new mothers is divided into two groups: smokers and non-smokers. Independent samples of 40 mothers are drawn from each group in order to make comparisons. One comparison made is birth weight. The babies of the 40 smokers average 112.1 ounces with a standard deviation of 16.1 ounces while those of the 40 non-smokers average 123.1 ounces with an SD of 15.8 ounces.

   (a) [**5 marks**] Is it clear that smokers have lower birth weight babies?

   (b) [**5 marks**] Give a 90 percent confidence interval for the difference in mean birth weights between smoking and non-smoking mothers.

4. The following sequence of questions concern the following model. Imagine we have 3 dice which have been carefully manufactured so that they all have exactly the same weight, θ. We begin by weighing 1 of the dice and recording Y_1 which you may assume has mean θ. The error in the measurement, namely $Y_1 - \theta$ has a normal distribution with mean 0 and standard deviation σ. In order to keep this problem simple you may assume that $\sigma = 1$ and that you know this somehow.

   Then you weigh 2 of the dice together and record Y_2 whose mean is 2θ. Assume that the error $Y_2 - 2\theta$ has a normal distribution with mean 0 and standard deviation σ.

   In this problem you are to compare two estimators for θ.

   (a) **[5 marks]** The first estimator is based on the idea that $Y_2/2$ has mean θ – the same as Y_1. This estimator is the average of these two.

   $$\hat{\theta}_1 = \frac{Y_1 + Y_2/2}{2}$$

   Find the bias, standard error and mean square error of $\hat{\theta}_1$.

(b) **[5 marks]** Another estimator is obtained by least squares. Derive the formula for the least squares estimate of θ; call this estimator $\hat{\theta}_2$.

(c) **[5 marks]** Find the bias, standard error and mean squared error of $\hat{\theta}_2$.

(d) **[2 marks]** Based on these calculations, which is the better estimator of θ, $\hat{\theta}_1$ or $\hat{\theta}_2$?

5. Suppose X has density

$$f(x,\theta) = \begin{cases} \frac{\theta}{x^2} & x > \theta \\ 0 & x \le \theta \end{cases}$$

(a) **[2 marks]** Find the cumulative distribution function of X.

(b) **[2 marks]** For any b such that $1 \le b$ find

$$P(1 \le \frac{X}{\theta} \le b)$$

and then find the values of b for which this probability is 0.025 and 0.975.

(c) **[5 marks]** Use the results of the previous problem to find a 95% confidence interval for θ.

(d) **[1 mark]** Evaluate your interval if we observe $X = 40$.

6. In the October 7, 2014 issue of the *Canadian Medical Association Journal* a randomized controlled double blind study of 378 patients studied the effect of melatonin on delirium. The treatment group had 186 patients and observed 55 cases of delirium while the control group had 192 patients and 49 cases of delirium.

   (a) **[1 mark]** If the treatment is completely ineffective what is the estimated probability of delirium in this group?

   (b) **[2 marks]** What is the estimated standard error for the estimate in the previous part.

   (c) **[5 marks]** Is there clear evidence of a difference in either direction in delirium rates between treatment and control?

7. I have regularly used the heights of 1078 father / adult son pairs gathered in Victorian England. We are interested in predicting the heights of sons from the heights of fathers supposing that the relationship is described by a straight line. Fathers average 67.69 inches in height with a standard deviation of 2.74 inches. Sons average 68.68 inches in height with a standard deviation of 2.81 inches. The average of the products (son's height times father's height) is 4652.895 square inches. Students writing in Burnaby or at the CSD get to see the following sentence which wass added after the exams were printed; students in Surrey will be shown the sentence on the board. The Error Sum of Squares $\sum_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$ is 6390.331 square inches.

   (a) **[2 marks]** Show that

$$\overline{xy} - \bar{x}\bar{y} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

   (b) **[3 marks]** Estimate the slope and intercept of the least squares line for predicting son's heights from father's heights.

(c) [**5 marks**]Give a 95% confidence interval for the true slope for the population these families are drawn from.

1a		5	5a		2
1b		5	5b		2
2		5	5c		5
3a		5	5d		1
3b		5	6a		1
4a		5	6b		2
4b		5	6c		5
4c		5	7a		2
4d		2	7b		3
			7c		5
			Total		70