Math Stat, Test 2, due April 15, noon

You may use the book, lecture notes, homework with solutions, and a calculator.

Give clear and complete solutions and state assumptions you are making.

1. A politician claims to have support from more than 50% of the population. In an opinion poll, 1166 people were asked and 612 of them supported the politician.

(a) Find a symmetric 95% confidence interval for the unknown proportion of supporters p. Is there support for the claim?
(b) If you want to do a new poll with a margin of error that is at most $\pm 1\%$ (± 0.01), how large should your sample be? Do it in two different ways.
(c) Instead of a confidence interval you may do a hypothesis test. State the relevant null hypothesis and alternative hypothesis (you do not have to do the actual test).

2(a) You are regularly monitoring a water pollutant at a measuring station, using confidence intervals for the mean μ. If it exceeds a certain critical level c, a warning is issued. New data shows that the pollutant is much more dangerous than previously thought and you don't want to risk that an increase in the pollutant goes unnoticed. Keeping the sample sizes constant, should you increase or decrease your confidence level? Give arguments.
(b) Make up an example where the adjustment of the confidence level should go in the opposite direction compared to **(a)**.

3. The web site realclearpolitics.com publishes opinion polls from a large number of different polling companies, news media, etc. For example, in March 2014 it published 19 polls regarding President Obama's approval ratings. Assume all polls have 95% confidence levels.

(a) What is the probability that a given poll does not capture the true approval rating within its margin of error?
(b) What is the probability that at least one of these polls does not capture the true approval rating within its margin of error?

4. True or false? Give arguments.

(a) If X and Y are independent $N(0,1)$ and $T = X/Y$, then T has a t distribution.
(b) If X has a a t_m distribution, then X^2 has an $F_{1,m}$ distribution.
(c) If X has an $F_{1,m}$ distribution, then $\sqrt{X}$ has a t_m distribution.
(d) If U and V are independent $N(0,1)$ and $T = U/|V|$, then T has a t distribution.
(e) If X and Y are independent $N(0,1)$ and $S = (X+Y)^2/2$, then S has a chi square distribution.

5. In problem 4(a) on HW 5, you considered the model $y = bx + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2)$ and showed that the MLE of b based on the observations $(x_1, Y_1), ..., (x_n, Y_n)$ is

$$\widehat{b} = \frac{S_{xY}}{S_{xx}}$$

where $S_{xx} = \sum_{k=1}^{n} x_k^2$ and $S_{xY} = \sum_{k=1}^{n} x_k Y_k$.

(a) Find the variance of $\widehat{b}$.
(b) Now consider the quadratic model $y = bx^2 + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2)$. Find the MLE of b based on the observations $(x_1, Y_1), ..., (x_n, Y_n)$.

6. Consider the confidence intervals for μ in a normal distribution when σ^2 is known and unknown, respectively, for the confidence level q. Let L_k (k for "known") and L_u be the lengths of the intervals.

(a) State expressions for L_k and L_u in terms of σ, s, n etc.
(b) Let F_{n-1} denote the cdf of the chi-square distribution with $n-1$ degrees of freedom. Express the probability that $L_u \leq L_k$ in terms of F_{n-1}.
(c) What happens to the probability in **(b)** as the sample size n increases (goes to infinity)? You may use the fact that the mean in a chi-square distribution equals the degrees of freedom, and what you know about sums of i.i.d. random variables. If you don't know how to do this, you can give an educated guess supported by an intuitive argument.