14.30 Introduction to Statistical Methods in Economics
Spring 2009

Problem Set #7

14.30 - Intro. to Statistical Methods in Economics

Instructor: Konrad Menzel

Due: Tuesday, April 14, 2009

Question One

We just learned about the standard Normal distribution with PDF $\phi(z)$ and CDF $\Phi(z)$. Let's familiarize ourselves with it, as we will be using it a lot in the future.

1. Find $u = \Phi(z)$ for $z = \{-3, -2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3\}$.

2. Find $u = 1 - Pr(|Z| \leq z)$ for $z = \{0, 0.5, 1, 1.5, 2, 2.5, 3\}$ using your answers from part (1). Explain (via math or words) how you obtained your answers. We call these values of u "p-values" which stands for "probability values" of a result at least that extreme occuring. Memorize these seven values–you will certainly use them in the future.

3. Find $z = \Phi^{-1}(u)$ for $u = \{0.001, 0.005, 0.01, 0.02, 0.05, 0.10, 0.20\}$.

4. Use the results from part (3) to obtain $z = \Phi^{-1}(1 - u)$ for each u.

5. Use the results and/or methods from part (3) and (4) to obtain z where $1 - Pr(|Z| \leq z) = u$ for each u in part (3). These values of z are called the "two-sided α-level critical values" where $\alpha = u$ in this example. For example, we say, "The (two-sided) 10% critical value of the standard normal distribution is $z = \ldots$" These will be useful in the future. Memorize these seven values as well.

6. Finally, interlace (i.e. sort) the values of z and u in a short table with the 14 values of z and u you obtained in parts (2) and (5). Start with $z = 0$ and $p = 1$ and put the z's in the right order with their corresponding probabilities.

Question Two

1. Given a random variable X, define the standardization Z of X and derive its variance.

2. Suppose $\bar{X}_{20}$ is the mean of a sample of $n = 20$ i.i.d. observations with $X_i \sim N(1, 2)$. What is the expected mean and variance of this average? What is its distribution (Hint: see lecture notes, around Proposition 24)? What is the probability that the sample mean $\bar{X}_{20}$ is between 0.75 and 1.25?

3. In Pset #5, we used the convolution formula to determine the finite distribution of the average (and sum) of k i.i.d. exponential random variables, noting that the mean did not depend on the distribution. What is the variance of the sum of k i.i.d. random variables? What's the variance of their average? We noticed (in the solution set to Pset #5) that the average over the exponential RVs approaches the normal distribution. Do you think that this property was specific to the exponential distribution?

Question Three

Suppose that a random sample of size n is taken from a normal distribution with mean μ and variance 3, and that the sample mean, $\bar{X}_n$, is calculated.

1. What does n need to be so that the probability of $\bar{X}_n$ being within 0.1 of μ is at least 90%?[1]

2. What does n need to be so that the probability of $\bar{X}_n$ being within 0.01 of μ is at least 90%?

3. In my experiments at Yahoo!, I have been looking at treatment and control differences to determine the effectiveness of online display advertising. The standard deviation of (scaled) weekly sales is R$15.00 and average weekly purchases are $R\$1.00$ for the control group and $R\$1.00 + \delta$ for the treatment group. If I am constrained to have only 25% of my observations in the control group and the remaining 75% in the treatment group, what is the variance of the difference between $\bar{X}_T - \bar{X}_C$ where $\bar{X}_T$ is the sample average over all treatment individuals and $\bar{X}_C$ is the average over all control individuals? How large does N have to be in order for the probability of $\bar{X}_T - \bar{X}_C > 0$ is at least 95%? Assume the sample averages for the treatment and control groups are normally distributed. Do this for general δ and then evaluate it for $\delta = R\$0.05$ and $\delta = R\$0.10$. Are these values of N large? Comment on the results.

Question Four

Use a table, calculator, internet, simulation, or any other method (besides cheating) to determine the following critical values:

1. $T \sim t$ distribution: $1 - P(|T| \leq t; dof) = 0.05$ for dof = 5, 10, 20, 30, 50. How do these compare to the 0.05 critical values of the Normal?

2. $X \sim \chi^2_k$ distribution: $P(X \geq x; k) = 0.05$ for $k = 1, 2, 3, 4, 5, 100$. How do these compare to the 0.05 critical values of the Normal? (Hint: Divide by k and take the

[1] These are what we call power calculations and help us craft surveys and experiments to guarantee before spending lots of money that we'll be able to detect economically meaningful effects, if they are present. For example, we usually think of log wages as being normally distributed. Often times we need to know how many observations we need in a survey for the average log wage to be close enough to the true mean. Further, as we will learn about the Central Limit Theorem, we'll realize that for any average that we might be interested, for $n \geq 30$, the distribution we sample from will generally not affect our calculation of n.

square root–kind of like an average. Focus on $k = 1$ and just compare the higher degrees of freedom to the critical values of $N(k, 2 \cdot k)$.) A few of these may be worth memorizing as well.

Question Five

What is the distribution of $\frac{Y_1/k_1}{Y_2/k_2}$ where $Y_1 \sim \chi^2_{k_1}$ and $Y_2 \sim \chi^2_{k_2}$? What happens to the value and variance of the denominator when when $k_2 \to \infty$? (Hint: What is the sum of two i.i.d. χ^2_1 random variables? Then, what's the variance of their average? And how about the sum and average of k_2 of them?) What distribution do you think that means that the ratio converges to (holding k_1 fixed)?