SIEO 3600: Final practice problems. Spring 2008.

1. (a) Suppose that T has a Student's $t-$ distribution with $v = 9$ degrees of freedom. Without using any $t-$ tables, etc., explain why it MUST hold that

$$P(|T| > 9/\sqrt{7}) \leq 1/9.$$

SOLUTION:
Since $\mu = E(T) = 0$, Chebychev's inequality yields $P(|T| > k\sigma) \leq 1/k^2$, for any integer k, where $\sigma^2 = Var(T)$. But (see Text, Page 269) $Var(T) = v/(v-2) = 9/7$, and thus $\sigma = 3/\sqrt{7}$; choosing $k = 3$ then yields the result.

(b) Suppose you only knew that $v \geq 9$. Explain why $P(|T| > 9/\sqrt{7}) \leq 1/9$ still remains valid.

SOLUTION:
$v/(v-2)$ is a decreasing (down to 1) function of $v \geq 3$; thus for $v \geq 9$, it follows that $v/(v-2) \leq 9/7$, or $3/\sqrt{7} \geq \sqrt{v/(v-2)} = \sigma$. Thus $P(|T| > 9/\sqrt{7}) = P(|T| > 3(3/\sqrt{7})) \leq P(|T| > 3\sigma) \leq 1/9$.

2. Suppose we want a continuous rv X that has density $f(x) = 3(1-x)^2, \ x \in (0,1)$.

(a) Show that $X = 1 - U^{1/3}$ works, where U denotes a random variable with a continuous uniform distribution over the interval $(0,1)$; $P(U \leq x) = x, \ x \in (0,1)$.

SOLUTION: With $h(u) = 1 - u^{1/3}$, The density of $X = 1 - U^{1/3} = h(U)$ is given by $f_X(x) = f_U(h^{-1}(x))|J(x)|$, where $J(x) = (h^{-1})'(x)$. Here $f_U(u) = 1, \ u \in (0,1)$ and $h^{-1}(x) = 1 - x^3$; $|J(x)| = 3(1-x)^2$. Thus, indeed, $f_Y(x) = 3(1-x)^2, \ x \in (0,1)$.

(b) Show that $X = \min\{U_1, U_2\}$ works, where U_1 and U_2 denote independent uniforms over the interval $(0,1)$.

SOLUTION:
Let $m = \min\{U_1, U_2\}$. We must show that m has density $f_m(x) = 3(1-x)^2, \ x \in (0,1)$. Observe that $\min\{U_1, U_2\} > x$ if and only if BOTH $U_1 > x$ and $U_2 > x$ since if the minimum is $> x$ then so is the maximum, so both must be $> x$. Using this fact: $F_m(x) = P(\max\{U_1, U_2\} \leq x) = 1 - P(\max\{U_1, U_2\} > x) = 1 - P(U_1 > x, \ U_2 > x) = 1 - P(U > x)^2$ (from independence of U_1 and U_2)
$= 1 - (1-x)^2$. Thus $f_m(x) = F'_m(x) = 3(1-x)^2, \ x \in (0,1)$.

3. 5% of new NYC taxis still do not have credit card capabilities (CC). Out of a random sample of 80 taxis, estimate the probability that at least 4 do not have CC.

SOLUTION:

Letting N denote the number out of 80 that do not have CC, N has a Binomial (n,p) distribution with $n = 80$ and $p = 0.05$. Since n is large and p is small, we can approximate by a Poisson rv Y with mean

$np = 4$. $P(X \geq 4) = 1 - P(X \leq 3) \approx 1 - P(Y \leq 3) =$
$1 - e^{-4}(1 + 4 + 4^2/2 + 4^3/3!) = 1 - e^{-4}(1 + 4^2/2 + 4^3/3!) = 1 - e^{-4}(341/3) = 0.371$.

4. A random sample of 20 glasses of tomato juice found that the mean salt content per glass was 244 milligrams with a sample standard deviation of 24.5 milligrams. Does this suggest at the 0.05 level of significance that the average salt content in a glass of tomato

juice is greater than 220 milligrams (which is the human requirement per day)? Assume that the distribution of salt content is approximately normal.

SOLUTION:

Under the null hypothesis $H_0 : \mu = 220$ (versus $H_1 : \mu > 220$), and due to the "approximately normal" assumption, we can treat $\frac{\overline{X}(20)-220}{s(20)/\sqrt{20}}$ as a $t-$ distribution with $n - 1 = 19$ degrees of freedom. Plugging in $\overline{X}(20) = 244$ and $s(n) = 24.5$ yields $t_0 = 4.38$. We are asked to do a one-sided test, so we need the value $t_{0.05,19} = 1.729$. We reject H_0 since $t_0 = 4.38 > 1.729$, and conclude that there is significant evidence that the average salt content in a glass of tomato juice is greater than 220 milligrams.

5. The City of New York is considering to use a new kind of traffic light at the busiest street intersections so as to reduce accidents in that area. 9 intersections were chosen, and measured for accidents per year both before installing the new light, and then again after. The following data gives the results. You are to make a one-sided test of the hypothesis that the new lights make no difference versus the new lights do reduce accidents (by comparing means). Use $\alpha = 0.05$, and assume that the distribution of accidents is approximately normal. What is your conclusion?

Intersection	Before	After
1	30	30
2	45	40
3	26	25
4	25	23
5	34	30
6	51	49
7	46	41
8	32	35
9	30	28

SOLUTION:

Here we use the "paired observations" method with $D_1 = 30 - 30 = 0$, $D_2 = 45 - 40 = 5$ and so on, up to $D_9 = 30 - 28 = 2$. $H_0 : \mu_D = d_0 = 0$ versus $H_1 : \mu_D > 0$. $n = 9$ is small and the data is said to be approximately normal so we use the T distribution with $v = 9 - 1 = 8$ degrees of freedom. The sample mean computes as $\overline{D}(9) = 2$, and the sample variance computes as $S^2(9) = 6.50$ so the sample standard deviation is $S(9) = 2.55$. $t_{0.05,8} = 1.86$. Our t value,

$$t_0 = \frac{\overline{D}(9) - d_0}{S/\sqrt{9}} = 2.4,$$

well into the critical region (e.g., $2.4 > 1.86$), so we reject H_0 in favor of H_1; we conclude that the new lights do reduce accidents.

6. For each of the following, tell wether the sample statistic to be used is the normal, the Student's t or the χ^2. (Justify your answers.)

   (a) A random sample of size $n = 140$ light bulbs are selected so as to estimate the mean lifetime of a bulb. You wish to construct a 95% confidence interval for the unknown mean.
   SOLUTION: Here, since n is quite large, we can use the normal (via the central limit theorem).

(b) A random sample of size $n = 15$ light bulbs are selected so as to estimate the mean lifetime of a bulb. You wish to construct a 95% confidence interval for the unknown mean. You know apriori that the lifetime distribution is very close to being normal.

SOLUTION: Here n is small, but we are told that the lifetime distribution is (approximately) normal. We do not know the real σ, so we will use the sample stdv $s(n)$; thus we will use the Student's t with $n - 1 = 14$ degrees of freedom.

(c) A random sample of size $n = 25$ light bulbs are selected so as to test the hypothesis that the number of defective bulbs (e.g., ones that blow out immediately) is Poisson or not.

SOLUTION: Classic application of the "goodness of fit test"; we use a χ^2.

(d) Given 20 pairs of data (x_i, y_i), $1 \leq i \leq 20$, you with to test if a linear regression model $Y = \beta_0 + \beta_1 x + \epsilon$ (with $\epsilon \sim N(0, \sigma^2)$) is appropriate by testing if $\beta_1 = 0$ or not.

SOLUTION: Because ϵ is normal, we know that in the end, our test involves a Student's t with $n - 2$ degrees of freedom. (Chapter 11, Page 405.)