**Problems 51-60 on Interval Estimation and Normal population (STAT2802 Statistical Models Tutorial notes for the week of 12-NOV-2012)**

51. To study the durability of a new paint for white center lines, a highway department painted test strips across heavily traveled roads in eight different locations, and electronic counters showed that they deteriorated after having been crossed by (to the nearest hundred) 142,600, 167,800, 136500, 108,300, 126,400, 133,700, 162,000, and 149,400 cars. Experts believe that the number of car crossings before deterioration has a normal distribution. Construct a 95% confidence interval for the average amount of traffic (car crossings) that this paint can withstand before it deteriorates.

52. Write down the p.d.f. of $\chi^2(\nu)$, the chi-square distribution with $\nu$ degrees of freedom. Verify that the p.d.f. integrates to 1. Find the MGF. Find the first 4 moments and the variance. (Hint: $\chi^2(\nu)$ density has kernel $x^{\nu/2}\, e^{-x/2}$.)

53. For large $n$, the sampling distribution of $S$ is sometimes approximated with a normal distribution having the mean $\sigma$ and the variance $\sigma^2/2n$. Why is this approximation valid? Also show that this approximation leads to the following $(1-\alpha)100\%$ large-sample confidence interval for $\sigma$:

$$\frac{s}{1+\frac{z_{\alpha/2}}{\sqrt{2n}}} < \sigma < \frac{s}{1-\frac{z_{\alpha/2}}{\sqrt{2n}}}.$$

54. If we want to construct a confidence interval for the mean, in what situation the normal distribution cannot help us and we have to resort to the *t*-distribution? Please explain concisely.

55. In the setting of comparing two normal samples <u>with equal variances</u>, show that

$$(\bar{X}_1-\bar{X}_2)-t\left(\frac{\alpha}{2},n_1+n_2-2\right)\cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2)+t\left(\frac{\alpha}{2},n_1+n_2-2\right)\cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

where $S_p=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$ is the pooled sample standard deviation and other symbols bear the usual meanings.

56. This question is about the <u>sample mean</u>. Population size is $N$. Sample size is $n$. Sampling is random and <u>without replacement</u>. Population mean is $\mu$, population sd is $\sigma$. A parenthetical superscript indicates which population the symbol is concerned, when more than one populations are involved. True or False or Filling the blank:

a) [T|F] $n\to\infty$. $\bar{X}\sim N(\mu,\sigma^2)$.

b) [T|F] $N<\infty$. The joint distribution of the sample is $\frac{1}{N^n}$.

c) [T|F] $N<\infty$. The <u>sample mean</u> is expected to be $\mu$ and has variance $\frac{1}{n}\sigma^2$.

d) $N<\infty$. The population consists of the first $N$ positive integers. Then the <u>sample sum</u> is expected to be $\frac{n(N+1)}{2}$ and has variance ____________.

e) $N\to\infty, n<\infty$. $\mathbb{P}(\mu-c<\bar{X}<\mu+c)\geq 1-$ ________, $c>0$. Hence, when $n\to\infty$, this probability is _____.

f) [T|F] $N^{(1)}\to\infty, N^{(2)}\to\infty$. Let $\bar{\delta}=\bar{X}^{(1)}-\bar{X}^{(2)}$. Then $\mathbb{E}[\bar{\delta}]=\mu^{(1)}-\mu^{(2)}$ and $\mathbb{V}[\bar{\delta}]=\frac{[\sigma^{(1)}]^2}{n^{(1)}}-\frac{[\sigma^{(2)}]^2}{n^{(2)}}$.

57. This question is about the sample standard deviation. Population size is $N$. Sample size is $n$. Sampling is random and without replacement. Population mean is $\mu$, population sd is $\sigma$. A parenthetical superscript indicates which population the symbol is concerned, when more than one populations are involved. True or False or Filling the blank:

a) [T|F] $N < \infty$. The population pairwise covariance is $\frac{\sigma^2}{N-1}$.

b) $N < \infty$. The population consists of the first $N$ positive integers. Then $\sigma^2 = \frac{1}{6}(N+1)(2N+1) - \mu^2$ and $S^2 =$ ______________________.

c) $N \to \infty$. $\mathrm{cov}(X_i - \bar{X}, \bar{X}) =$ _____.

d) [T|F] $N \to \infty$. $S^2 = \frac{n(\sum_{i=1}^n X_i^2) - (\sum_{i=1}^n X_i)^2}{n(n-1)}$.

e) $N \to \infty$. If the population is Normal, then $\bar{X}$ ___ $S^2$.

58. This question concerns the three distribution related to the Normal model: $\chi^2(df)$, $t(df)$, and $F\left(df^{(1)}, df^{(2)}\right)$. True or False or Filling the blank:

a) $Z \sim N(0,1) \Longrightarrow Z^2 \sim$_____ with mean ______ and variance _______.

b) $Z_1, \ldots, Z_n \overset{iid}{\sim} N(0,1) \Rightarrow \sum_{i=1}^n Z_i^2 \sim$ _____ with mean _______ and variance _______.

c) $\chi_1 \sim \chi^2(\nu_1)$ and $\chi_2 \sim \chi^2(\nu_2)$. If $\chi_1 \perp\!\!\!\perp \chi_2$ then $\chi_1 + \chi_2 \sim$ ________ with mean _______ and variance ________.

d) Population is normal. $f(n, S, \sigma) =$ ___________ ~ $\chi^2$(______).

e) Population is normal. $f(n, \bar{X}, S, \mu) =$ ____________ ~ $t$(______).

f) [T|F] The mean of the (centralized) $t$-distribution is positive, which is also its symmetry axis.

g) Populations are normal. $f(S_1, S_2, \sigma_1, \sigma_2) =$ ____________ ~ $F$(___________).

h) [T|F] The $F$-distribution is symmetric about 0.

59. Show that for $\nu > 2$, the variance of the $t$-distribution with $\nu$ degrees of freedom is $\frac{\nu}{\nu-2}$. *Hint: Make the substitution* $1 + \frac{x^2}{\nu} = \frac{1}{u}$. *The* $t(\nu)$ *density has kernel* $\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$. *Beta function:* $\int_0^1 u^{x-1}(1-u)^{y-1} du = B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

60. Show that for $\nu_2 > 2$, the mean of the $F$ distribution is $\frac{\nu_2}{\nu_2 - 2}$. *Hint: the* $F(\nu_1, \nu_2)$ *density is* $\frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2}-1} \left(1 + \frac{\nu_1}{\nu_2} x\right)^{-\frac{1}{2}(\nu_1+\nu_2)} \mathbb{I}(x > 0)$.