Solutions to Problems 51-60

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.

Solution. The population is normal. The sample size is 8. The sample mean is (142600+167800+136500+108300+126400+133700+162000+149400)/8=140838. The sample sd is sqrt((142600^2 + 167800^2 + 136500^2 + 108300^2 + 126400^2 + 133700^2 + 162000^2 + 149400^2- 8 * 140838^2) / (8 - 1)) = 19228.47. The 0.975 quantile of the t distribution with 8-1=7 degrees of freedom is 2.365. Therefore, the 95% CI for the population mean is [140838 ± 2.365 * 19228.47 / sqrt(8)] = [124762.10, 156912.90].

52. Write down the p.d.f. of $\chi^2(\nu)$, the chi-square distribution with ν 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 $x^{\nu/2}\,e^{-x/2}$.)

Solution. P.d.f of $\chi^2(\nu)$ is $f_X(x)=\frac{1}{c}x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}\mathbb{I}(x>0)$. The normalizing constant is $c=\int_0^\infty x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}dx=2^{\frac{\nu}{2}}\int_0^\infty y^{\frac{\nu}{2}-1}e^{-y}dy=2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)$.

$$\varphi_X(t)=\mathbb{E}(e^{tX})=\frac{1}{c}\int_0^\infty e^{tx}x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}dx=\frac{1}{c}\int_0^\infty e^{-(1-2t)\frac{x}{2}}x^{\frac{\nu}{2}-1}dx=\frac{2^{\frac{\nu}{2}}}{(1-2t)^{\frac{\nu}{2}}}\frac{1}{c}\Gamma(\frac{\nu}{2})=(1-2t)^{-\frac{\nu}{2}}$$

$$\mathbb{E}(X)=\frac{1}{c}\int_0^\infty x^{\frac{\nu}{2}}e^{-\frac{x}{2}}dx=\frac{2^{\frac{\nu}{2}+1}\Gamma(\frac{\nu}{2}+1)}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}=\nu$$

$$\mathbb{E}(X^2)=\frac{1}{c}\int_0^\infty x^{\frac{\nu}{2}+1}e^{-\frac{x}{2}}dx=\frac{2^{\frac{\nu}{2}+2}\Gamma(\frac{\nu}{2}+2)}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}=2^2(\frac{\nu}{2}+1)\frac{\nu}{2}=(\nu+2)\nu$$

$$\mathbb{E}(X^3)=\frac{1}{c}\int_0^\infty x^{\frac{\nu}{2}+2}e^{-\frac{x}{2}}dx=\frac{2^{\frac{\nu}{2}+3}\Gamma(\frac{\nu}{2}+3)}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}=2^3(\frac{\nu}{2}+2)(\frac{\nu}{2}+1)\frac{\nu}{2}=(\nu+4)(\nu+2)\nu$$

$$\mathbb{E}(X^4)=(\nu+6)(\nu+4)(\nu+2)\nu$$

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 2\nu$$

53. For large n, the sampling distribution of S is sometimes approximated with a normal distribution having the mean σ 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 σ:

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

Solution. Bear in mind that $n \to \infty$ here. Firstly, $\frac{(n-1)S^2}{\sigma^2} \to \chi^2(n-1)$ is well known. **Lemma**. If $X \sim \chi^2(\nu \to \infty)$ then $\sqrt{2X} - \sqrt{2\nu} \to N(0,1)$. **Proof of Lemma**. $\mathbb{P}\left(\sqrt{2X} - \sqrt{2\nu} \le t\right) = \mathbb{P}\left(\frac{X-\nu}{\sqrt{2\nu}} \le \frac{t^2}{\sqrt{2\nu}} + t\right) \xrightarrow{\because X \to N(\nu, 2\nu)} \Phi\left(t + \frac{t^2}{\sqrt{2\nu}}\right) \to \Phi(t)$. **End of Pf of Lemma**. Hence $S \to \frac{\sigma}{\sqrt{2(n-1)}}\sqrt{2\chi^2(n-1)} \to \frac{\sigma}{\sqrt{2(n-1)}} N\left(\sqrt{2(n-1)}, 1\right) \to N\left(\sigma, \frac{\sigma^2}{2n}\right)$. Thus, we solve the inequality $-z_{\alpha/2} < \frac{S-\sigma}{\sigma/\sqrt{2n}} < z_{\alpha/2}$ to get the $(1-\alpha)100\%$ CI for σ.

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.

Solution. For large sample size n, firstly, by Central Limit Theorem, the sample mean is normal: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. If the variance parameter σ^2 is known, then CI for μ can be directly obtained from this sampling distribution of $\bar{X}$. But most often, the variance parameter σ^2 is unknown, we have to find a pivot whose pivot distribution does not involve σ^2. And it is found that $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.

55. In the setting of comparing two normal samples with equal variances, 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.

Solution. It is sufficient to show that $T = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ has a t-distribution with n_1+n_2-2 degrees of freedom. Since the two equi-variance populations are $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, therefore $Z = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}} \sim N(0,1)$. On the other hand, $Y = \frac{(n_1+n_2-2)S_p^2}{\sigma^2} = \frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_1+n_2-1)$ as two independent chi-squares add to another chi-square. Thus $\frac{Z}{\sqrt{\frac{Y}{n_1+n_2-2}}} \sim t(n_1+n_2-2)$.

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

Solution. (correct statements)

a) `[F] Correct:` $n \to \infty$. $\bar{X} \sim N(\mu, \sigma^2/n)$.

b) `[F] Correct:` $N < \infty$. The joint distribution of the sample is $\frac{1}{N(N-1)\cdots(N-n+1)}$.

c) `[F] Correct:` $N < \infty$. The sample mean is expected to be μ and has variance $\frac{1}{n}\sigma^2 \cdot \frac{N-n}{N-1}$.

d) $N < \infty$. The population consists of the first N positive integers. Then the sample sum is expected to be $\frac{n(N+1)}{2}$ and has variance $\frac{1}{12}n(N+1)(N-n)$.

e) $N \to \infty, n < \infty$. $\mathbb{P}(\mu - c < \bar{X} < \mu + c) \geq 1 - \frac{\sigma^2}{nc^2}$, $c > 0$. Hence, when $n \to \infty$, this probability is 1.

f) `[F] Correct:` $N^{(1)} \to \infty, N^{(2)} \to \infty$. 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 μ, population sd is σ. 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) `[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 $S^2 = \frac{\sum_{i=1}^n X_i^2}{n-1} - \frac{n\bar{X}^2}{n-1}$.

c) $N \to \infty$. $\mathrm{cov}(X_i - \bar{X}, \bar{X}) = 0$.

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

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

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

Solution. (correct statements)

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

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

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

d) Population is normal. $f(n,S,\sigma)=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$.

e) Population is normal. $f(n,\bar{X},S,\mu)=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$.

f) `[F] Correct:` The mean of the (centralized) t-distribution is 0, which is also its symmetry axis.

g) Populations are normal. $f(S_1,S_2,\sigma_1,\sigma_2)=\frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}\sim F(n_1-1,n_2-1)$.

h) `[F] Correct:` The F-distribution is not symmetric and is distributed on the positive half line $[0,+\infty)$.

59. Show that for $\nu>2$, the variance of the t-distribution with ν 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:* $\int_0^1 u^{x-1}(1-u)^{y-1}du=B(x,y)=\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

Solution. The (central) t-distribution has a symmetric density about 0. For df=ν >1, it has an expectation equal to 0. The second moment exists when $\nu>2$. To find it we try to fit the integral into a Beta function. With $u=\left(1+\frac{x^2}{\nu}\right)^{-1}\in(0,1]\Rightarrow x=\sqrt{\nu}\sqrt{\frac{1}{u}-1}$ and $\frac{du}{dx}=-\frac{2xu^2}{\nu}\Leftrightarrow dx=-\frac{\sqrt{\nu}}{2u^2}\left(\frac{1}{u}-1\right)^{-\frac{1}{2}}du$,

$$\mathbb{V}(X)=\mathbb{E}(X^2)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\int_{-\infty}^{+\infty}x^2\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}dx=\frac{\Gamma\left(\frac{\nu+1}{2}\right)\cdot\nu}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}}\int_{u=0}^{1}(1-u)^{\frac{1}{2}}u^{\frac{\nu}{2}-2}du=\frac{\Gamma\left(\frac{\nu+1}{2}\right)\cdot\nu}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}}B\left(\frac{3}{2},\frac{\nu}{2}-1\right)=\frac{\nu}{\nu-2}$$

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* $\frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}x^{\frac{\nu_1}{2}-1}\left(1+\frac{\nu_1}{\nu_2}x\right)^{-\frac{1}{2}(\nu_1+\nu_2)}\mathbb{I}(x>0)$.

Solution. With $u=(1+\frac{\nu_1}{\nu_2}x)^{-1}\Rightarrow x=\left(\frac{1}{u}-1\right)\frac{\nu_2}{\nu_1}$ and $dx=-u^{-2}\frac{\nu_2}{\nu_1}du$:

$$\mathbb{E}(X)=\frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}\int_0^{\infty}x^{\frac{\nu_1}{2}}\left(1+\frac{\nu_1}{\nu_2}x\right)^{-\frac{1}{2}(\nu_1+\nu_2)}dx=\frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\frac{\nu_2}{\nu_1}\int_0^1(1-u)^{\frac{\nu_1}{2}}u^{\frac{1}{2}\nu_2-2}du=\frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\frac{\nu_2}{\nu_1}B\left(\frac{\nu_1}{2}+1,\frac{\nu_2}{2}-1\right)=\frac{\nu_2}{\nu_2-2}$$