STA 580 — Fall 2008 — Dr. Charnigo

Written Assignment 2 Solutions

1a. From Written Assignment 1 we know that $\bar{x} = 138.72$, $s = 15.29$, and $n = 25$. The $100(1-\alpha)\%$ "small sample" confidence interval for μ is

$$\bar{x} \pm t_{n-1,1-\alpha/2} s/\sqrt{n}.$$

Putting $\alpha = 0.05$ and noting that $t_{n-1,1-\alpha/2} = t_{24,0.975} = 2.064$, we obtain

$$138.72 \pm 6.31, \text{ which is } 132.41 \text{ to } 145.03.$$

The systolic blood pressure measurements must be normally distributed to ensure the validity of this confidence interval, as a normal population is required whenever we construct a "small sample" confidence interval for μ using the formula

$$\bar{x} \pm t_{n-1,1-\alpha/2} s/\sqrt{n}.$$

1b. Since 140 is contained in the 95% confidence interval, 140 is a plausible value for μ. So we do not anticipate rejecting $H_0 : \mu = 140$ in favor of $H_1 : \mu \neq 140$. Formally, we conduct a level α "small sample" test of $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ by constructing the test statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

and comparing its absolute value to $t_{n-1,1-\alpha/2}$. In this case, with $\mu_0 = 140$ and $\alpha = 0.05$, we have

$$t = \frac{138.72 - 140}{15.29/\sqrt{25}} = \frac{-1.28}{3.06} = -0.418,$$

whose absolute value is less than $t_{24,0.975} = 2.064$. Therefore we do not reject $H_0 : \mu = 140$ in favor of $H_1 : \mu \neq 140$.

1c. The general formula for (approximate) power in testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ at level α is

$$\Phi\left(-z_{1-\alpha/2} + \frac{|\mu_0 - \mu_1|\sqrt{n}}{\sigma}\right).$$

We have $\alpha = 0.05$, so that $-z_{1-\alpha/2} = -z_{0.975} = -1.960$. We also have $\mu_0 = 140$ and $n = 75$. Taking $\mu_1 = 138.72$ and $\sigma = 15.29$, as we have no compelling reason to do otherwise, we find that the power is

$$\Phi\left(-1.960 + \frac{|140 - 138.72|\sqrt{75}}{15.29}\right) = \Phi(-1.235) = 10.8\%.$$

Remark. The 10.8% is quite small but intuitively plausible. If μ really were 138.72, then the null hypothesis would almost be true. Rejecting a null hypothesis that is almost true requires an immense amount of data, and $n = 75$ is not an immense sample size.

1d. The general formula for (approximate) sample size in testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ at level α, when one desires power $1 - \beta$, is

$$n = \frac{\sigma^2 (z_{1-\beta} + z_{1-\alpha/2})^2}{(\mu_0 - \mu_1)^2}.$$

With $1-\beta = 0.90$ we have $z_{1-\beta} = z_{0.90} = 1.282$. So the required sample size is

$$\frac{15.29^2(1.282+1.960)^2}{(140-138.72)^2} \approx 1500.$$

1e. The $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}} \text{ to } \frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}.$$

Putting $\alpha = 0.05$ and noting that $\chi^2_{n-1,1-\alpha/2} = \chi^2_{24,0.975} = 39.36$, $\chi^2_{n-1,\alpha/2} = \chi^2_{24,0.025} = 12.40$, we obtain

$$\frac{(24)15.29^2}{39.36} \text{ to } \frac{(24)15.29^2}{12.40}, \text{ which is } 142.6 \text{ to } 452.5.$$

Since 142.6 to 452.5 is a range of plausible values for σ^2, $\sqrt{142.6} = 11.94$ to $\sqrt{452.5} = 21.27$ is a range of plausible values for $\sqrt{\sigma^2} = \sigma$. The systolic blood pressure measurements must be normally distributed to ensure the validity of this confidence interval, as a normal population is required whenever we construct a confidence interval for σ^2 using the formula

$$\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}} \text{ to } \frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}.$$

2a. We can easily calculate that $\bar{x} = 127.52$, $s = 30.95$, and $n = 200$. The $100(1-\alpha)\%$ "large sample" confidence interval for μ is

$$\bar{x} \pm z_{1-\alpha/2} s/\sqrt{n}.$$

Putting $\alpha = 0.05$ and noting that $z_{1-\alpha/2} = z_{0.975} = 1.960$, we obtain

$$127.52 \pm 4.29, \text{ which is } 123.23 \text{ to } 131.81.$$

The plasma glucose concentration measurements need not be normally distributed to ensure the validity of this confidence interval, as a normal population is not required to construct a "large sample" confidence interval for μ using the formula

$$\bar{x} \pm z_{1-\alpha/2} s/\sqrt{n}.$$

Remark 1. Do not confuse approximate normality of (random) $\bar{X}$, which justifies this confidence interval, with normality of the population. The Central Limit Theorem ensures that, for large n, $\bar{X}$ is approximately normal even if the population is not normal.

Remark 2. If one uses SAS to obtain a confidence interval, then the plasma glucose concentrations must be normally distributed. This is because SAS calculates a "small sample" confidence interval — i.e., uses $t_{n-1,1-\alpha/2}$ instead of $z_{1-\alpha/2}$ — regardless of n, and a "small sample" confidence interval requires a normal population.

2b. We conduct a level α "large sample" test of $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ by constructing the test statistic

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

and comparing it to $z_{1-\alpha}$. In this case, with $\mu_0 = 126$ and $\alpha = 0.05$, we have

$$z = \frac{127.52 - 126}{30.95/\sqrt{200}} = \frac{1.52}{2.19} = 0.694,$$

which is less than $z_{0.95} = 1.645$. Therefore we do not reject $H_0 : \mu = 126$ in favor of $H_1 : \mu > 126$.

2c. The general formula for power in testing $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ at level α is

$$\Phi\left(-z_{1-\alpha} + \frac{|\mu_0 - \mu_1|\sqrt{n}}{\sigma}\right).$$

[Note that μ_1 must be larger than μ_0 to apply this formula, for otherwise $H_1 : \mu > \mu_0$ would not be true.] We have $\alpha = 0.05$, so that $-z_{1-\alpha} = -z_{0.95} = -1.645$. We also have $\mu_0 = 126$ and $n = 500$. Taking $\mu_1 = 127.52$ and $\sigma = 30.95$, as we have no compelling reason to do otherwise, we find that the power is

$$\Phi\left(-1.645 + \frac{|126 - 127.52|\sqrt{500}}{30.95}\right) = \Phi\left(-0.547\right) = 29.2\%.$$

2d. The general formula for sample size in testing $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ at level α, when one desires power $1 - \beta$, is

$$n = \frac{\sigma^2(z_{1-\beta} + z_{1-\alpha})^2}{(\mu_0 - \mu_1)^2}.$$

[Note that μ_1 must be larger than μ_0 to apply this formula, for otherwise $H_1 : \mu > \mu_0$ would not be true.] With $1 - \beta = 0.80$ we have $z_{1-\beta} = z_{0.80} = 0.842$. So the required sample size is

$$\frac{30.95^2(0.842 + 1.645)^2}{(126 - 127.52)^2} \approx 2565.$$

2e. We can easily calculate that $\hat{p} = 93/200 = 0.465$ and $n = 200$. The $100(1 - \alpha)\%$ "large sample" confidence interval for p is

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\hat{p}(1 - \hat{p})/n}.$$

Putting $\alpha = 0.05$ and noting that $z_{1-\alpha/2} = z_{0.975} = 1.960$, we obtain

$$0.465 \pm 0.069, \text{ which is } 0.396 \text{ to } 0.534.$$

The plasma glucose concentration measurements need not be normally distributed to ensure the validity of this confidence interval, as a normal population is not required to construct a "large sample" confidence interval for p using the formula

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\hat{p}(1 - \hat{p})/n}.$$

Remark. Do not confuse approximate normality of (random) $\hat{P}$, which justifies this confidence interval, with normality of the population. The Central Limit Theorem ensures that, for large n, $\hat{P}$ is approximately normal even if the population is not normal. Moreover, in constructing a confidence interval for p we are not really using the plasma glucose concentration measurements anyway; all we are paying attention to is whether someone is a "success" (plasma glucose concentration greater than 126) or a "failure" (plasma glucose concentration less than or equal to 126).