STAT 6200 Hwk 4 Solution

<u>Ch. 8</u> 8, 10, 12, 14 Ch. 9: 4, 5, 6, 8, 11, 13

8. $\mu = 29.5$, $\sigma = 9.25$ $n = 20$

a) $E(\bar{X}) = \mu$ so the mean of the sample means is the pop. mean, $\mu = 29.5$

b.) $Var(\bar{X}) = \frac{\sigma^2}{n}$, $SD(\bar{X}) = \sqrt{Var(\bar{X})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$ which is also called the standard error of $\bar{X}$.

c) The s.d. of the $\bar{X}$'s is $\frac{\sigma}{\sqrt{n}}$ where σ is the s.d. of the original measurements. So when the sample size is $n = 20$, the SD of the sample means is $\left(\frac{1}{\sqrt{20}}\right)^{th}$ or $\left(\frac{1}{4.472}\right)^{th}$ as large as the SD of the albumin levels.

d.) ~~Although~~ Since we don't know that the albumin levels are normally distributed, the $\bar{X}$'s aren't necessarily exactly normally distributed (they would be, if albumin was). However, because albumin is symmetric and the sample size is moderately large ($n=20$), the ~~CLT~~ normal dist'n should provide a good approximation to the dist'n of the $\bar{X}$'s by the CLT. That is, a histogram of the $\bar{X}$'s should be roughly normal, or bell-shaped.

e) According to the CLT, $\overline{X} \mathrel{\dot\sim} N(\mu, \sigma^2/n)$, so

$$P(\overline{X} > 33) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} > \frac{33-\mu}{\sigma/\sqrt{n}}\right) \approx P\left(Z > \frac{33-29.5}{9.25/\sqrt{20}}\right)$$

$$= P(Z > 1.69) = .046 \text{ or } 4.6\%$$

f.) $$P(\overline{X} < 28) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{28-\mu}{\sigma/\sqrt{n}}\right) \approx P\left(Z < \frac{28-29.5}{9.25/\sqrt{20}}\right)$$

$$= P(Z < -.73) = P(Z > .73) = .233 \text{ or } 23.3\%$$

g) $$P(29 < \overline{X} < 31) = P\left(\frac{29-\mu}{\sigma/\sqrt{20}} < \frac{\overline{X}-\mu}{\sigma/\sqrt{20}} < \frac{31-\mu}{\sigma/\sqrt{n}}\right)$$

$$\approx P\left(\frac{29-29.5}{9.25/\sqrt{20}} < Z < \frac{31-29.5}{9.25/\sqrt{20}}\right) = P(-.24 < Z < .75)$$

$$= P(Z < .75) - P(Z < -.24) = 1 - P(Z > .75) - P(Z > .24)$$

$$= 1 - .227 - .405 = .368 \text{ or } 36.8\%$$

10. $n = 40$ is a fairly large sample size, so the CLT should give a good approximation to the dist'n of $\overline{X}$, even though $\overline{X}$ was computed from non-normal data.

$$\overline{X} \mathrel{\dot\sim} N(\mu, \sigma^2/n) = N\left(1.81, \frac{(2.25)^2}{40}\right) = N(1.81, .127)$$

$\overline{X}$ should be approximately normal centered at $\mu = 1.81$ w/ S.D $= \sqrt{.127} = .356$

12. $\mu = 13.3, \ \sigma = 1.12$

a) Without knowing the distribution of hemoglobin, its impossible to know the exact distribution of the sample means. However, assuming that $n = 15$ is large enough for the CLT to hold,

$$\bar{X} \,\dot\sim\, N(\mu, \sigma^2/n) \text{ where } \mu = 13.3,\ \sigma = 1.12$$
and $n = 15$

$$P(13.0 < \bar{X} < 13.6) = P\left(\frac{13-\mu}{\sigma/\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{13.6-\mu}{\sigma/\sqrt{n}}\right)$$

$$\approx P\left(\frac{13-13.3}{1.12/\sqrt{15}} < Z < \frac{13.6-13.3}{1.12/\sqrt{15}}\right) = P(-1.04 < Z < 1.04)$$

$$= 1 - 2P(Z > 1.04) = 1 - 2(.149) = .702 \quad (70.2\%)$$

b) $$P(13.0 < \bar{X} < 13.6) \approx P\left(\frac{13-13.3}{1.12/\sqrt{30}} < Z < \frac{13.6-13.3}{1.12/\sqrt{30}}\right)$$

$$= P(-1.47 < Z < 1.47) = 1 - 2P(Z > 1.47) = 1 - 2(.071) = .858$$
(85.8%)

c) ~~Need $P\left(\frac{13.1-13.3}{}\right.$ $P(13.1 < \bar{X} < 13.5) = .95$~~

~~$\Rightarrow .95 = P\left(\frac{13.1-13.3}{}\right.$~~

Need $$.95 = P(13.1 < \bar{X} < 13.5) \approx P\left(\frac{13.1-\mu}{\sigma/\sqrt{n}} < Z < \frac{13.5-\mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left(\frac{13.1-13.3}{1.12/\sqrt{n}} < Z < \frac{13.5-13.3}{1.12/\sqrt{n}}\right) = 1 - 2P\left(Z > \frac{13.5-13.3}{1.12/\sqrt{n}}\right)$$

$= 1 - 2P\left(Z > \frac{(.2)\sqrt{n}}{1.12}\right) = 1 - 2P(Z > .1786\sqrt{n})$

$\Rightarrow .95 = 1 - 2P(Z > .1786\sqrt{n}) \Rightarrow 2P(Z > .1786\sqrt{n}) = 1 - .95 = .05$

$\Rightarrow P(Z > .1786\sqrt{n}) = \frac{.05}{2} = .025$

$\Rightarrow .1786\sqrt{n}$ should be the $.975^{th}$ percentile of the standard normal distribution

That is, $z_{.975} = 1.96 = .1786\sqrt{n}$

$\Rightarrow \sqrt{n} = \frac{1.96}{.1786} \Rightarrow n = \left(\frac{1.96}{.1786}\right)^2 = 120.4$ or

about $n = 121$ subjects.

d.) Similar to part c), we need $.95 = 1 - 2P\left(Z > \frac{13.4 - 13.3}{1.12/\sqrt{n}}\right)$

$\Rightarrow .95 = 1 - 2P(Z > .0893\sqrt{n}) \Rightarrow z_{.975} = 1.96 = .0893\sqrt{n}$

$\Rightarrow n = \left(\frac{1.96}{.0893}\right)^2 = 481.7 \approx 482$

14. $\mu = 172.2$, $\sigma = 29.8$

a) Should be approx'ly normal too, with mean $\mu = 172.2$ and

s.d. $= \frac{\sigma}{\sqrt{n}} = \frac{29.8}{\sqrt{25}} = 5.96$

14 b.) Want $\bar{X}_{.90}$. $\bar{X}_{.90} = \mu + z_{.90}(\sigma/\sqrt{n})$

$$= 172.2 + (1.282)(29.8/\sqrt{25}) = 179.8$$

c) Want $\bar{X}_{.20} = \mu + z_{.20}(\sigma/\sqrt{n}) = 172.2 + (-.8416)(29.8/\sqrt{25})$
$= 167.2$

d.) $P(\bar{X} \geq 190) \approx P\left(Z \geq \frac{190-\mu}{\sigma/\sqrt{n}}\right) = P\left(Z \geq \frac{190-172.2}{29.8/\sqrt{25}}\right)$

$= P(Z \geq 2.99) = .001$. Result is very unlikely if $\mu = 172.2$ and $\sigma = 29.8$, so its tempting to conclude that the population that I've drawn from has different values of these parameters (larger mean, most likely, than 172.2 lbs)

Ch. 9

4. Both the t and $N(0,1)$ distributions are parametric distributions centered at 0. They are also both bell-shaped. However, the t distribution is more dispersed than the $N(0,1)$. Also, the t has a parameter, its degrees of freedom, that quantifies how much more dispersed it is than the $N(0,1)$.
The $N(0,1)$ is appropriate if σ, the population s.d., is known. If σ is unknown, use the $t(n-1)$ dist'n.

5. $\sigma_s = 11.8$, $\sigma_d = 9.1$

a) $\bar{X}_s = 130$ 95% CI for μ_s is given by

$$\bar{X}_s \pm \underbrace{Z_{1-.05/2}}_{=Z_{.975}=1.96} \sigma/\sqrt{n} = 130 \pm 1.96(11.8/\sqrt{10})$$

$$= (122.7, 137.3)$$

b.) We can be 95% confident that μ_s, the mean systolic b.p among 30-34 female diabetes, is covered by (lies between) ~~122.7 and 137.3~~ the interval (122.7, 137.3). Or, if we were to repeatedly draw random samples of size 10 from this population, and compute 95% CI's for μ_s in this way, 95% of them would cover the true value of μ_s.

c.) $\bar{X}_d = 84$ 90% CI for μ_d is given by

$$\bar{X}_d \pm Z_{1-\frac{.10}{2}} \sigma/\sqrt{n} = 84 \pm 1.645(9.1/\sqrt{10})$$

$$= (79.27, 88.73)$$

d.) $$\bar{X}_d \pm \underbrace{Z_{1-\frac{.01}{2}}}_{=Z_{.995}=2.576} \sigma/\sqrt{n} = 84 \pm 2.576(9.1/\sqrt{10}) = (76.59, 91.41)$$

e) 99% CI is wider than 90% CI

6a) $P(t(5) > 2.015) = .05$ (can be obtained by looking up 2.015 in Table A.4, or from a computer program such as Minitab)

b.) $P(t(5) < -3.365) = P(t(5) > 3.365) = .01$ (Table A.4)

c) $P(-4.032 < t(5) < 4.032) = 1 - 2P(t(5) > 4.032)$
$= 1 - 2(.005) = .99$

d) $t_{.975}(5) = 2.571$

8. $n = 12$, $\bar{X}_1 = 4.49$ (FVC) $s_1 = .83$
$\bar{X}_2 = 3.71$, $S_2 = .62$ (FEV_1)

a) 95% CI for μ_1 given by

$$\bar{X}_1 \pm t_{1-\frac{.05}{2}}(n-1)\, s_1/\sqrt{n} = 4.49 \pm \underbrace{t_{.975}(11)}_{= 2.201} \frac{.83}{\sqrt{12}}$$
$$= (3.96, 5.02)$$

b.) $\bar{X}_1 \pm \underbrace{t_{1-\frac{.10}{2}}(11)}_{= 1.796} \frac{.83}{\sqrt{12}} = (4.06, 4.92)$ get's narrower.

c) $\bar{X}_2 \pm t_{1-\frac{.05}{2}}(11) \frac{s_2}{\sqrt{12}} = 3.71 \pm 2.201 \frac{.62}{\sqrt{12}} = (3.32, 4.10)$

d.) That they are normal

11. $\bar{X}_c = 3.143$, ~~S~~ $S_c = .5107$, $\bar{X}_a = \cancel{40.38}\ 40.375$, $S_a = 3.021$

a) 95% lowerbound given by

$$\bar{X}_c - t_{1-\underbrace{.05}_{=\alpha}}(n-1)\frac{S_c}{\sqrt{n}} = 3.143 - \underbrace{t_{.95}(7)}_{=1.895}\frac{.5107}{\sqrt{8}}$$

$$= 3.143 - 1.895(.5107)/\sqrt{8}$$

$$= 2.800$$

b.) $\bar{X}_a - 1.895(3.021)/\sqrt{8} = 38.35$

c) We can ~~by~~ be 95% confident that μ_a is not lower than 38.35, ~~so the~~ and $\bar{X}_a = 40.375$. Both of these values are w/in the normal range, so albumin seems normal.
However, we can be 95% confident that μ_c is above 2.800, which falls above the normal range for calcium. It appears they have elevated calcium.

13. In Minitab open the worksheet lowbwt. Then select

Data → Unstack columns... ~~.~~

Select the variable sbp into "Unstack the data in:"
" " " sex into "Using subscripts in:"
Select "Store unstacked data: After last column in use"
and check "Name the columns containing the unstacked data."
Hit OK

a) Then select
+b) Stat → Basic Statistics → 1-Sample t...

Select the variables 'sbp_0' and 'sbp_1' under "Samples in columns:" and then hit O.K.

The 95% CIs are (43.48, 49.45) for females
and (44.27, 51.45) for males

c) Since these CI's overlap substantially, we can't be sure that there's a difference between the pop. mean Sbp's for females & males