## MATH 156.3 — TEST III SOLUTIONS

**1** The Central Limit Theorem (CLT) says that
for data from an SRS of size $n$ (problem says SRS and that $n = 500$)
out of a population where
the population proportion of those who say yes to some question (here the question is "do you perform regular volunteer service?") is $\pi$ (problem says $\pi = .25$),
we have
the sample proportion $\hat{p}$ is approximately Normal with mean $\pi$ and standard deviation $\sqrt{\frac{\pi(1-\pi)}{n}}$, *i.e.*,

$$\hat{p} \text{ is } N\left(\pi,\ \sqrt{\frac{\pi(1-\pi)}{n}}\right) = N(.25,\ .01936)$$

So $P(\hat{p} > .282) = \texttt{normalcdf}(.282,\ 9999,\ .25,\ .01936) = .0513.$
and this is the probability that the sample proportion will be greater than 28.2%.

**2** **(a)** Let
$Y$ be the RV "distance to the sun"
$X$ be the RV "planet number, counting from Mercury = 1"
for
the OUs "planets"
out of
the population "planets in the Solar System."
The LSRL has equation

$$\hat{y} = a + bx \text{ where } b = r\frac{S_Y}{S_X} \text{ and } a = \overline{Y} - b\overline{X}.$$

Here
$r$ is the correlation coefficient of $X$ and $Y$
$S_X$ (respectively, $S_Y$) is the standard deviation of the $X$ data (resp., the $Y$ data)
$\overline{X}$ (resp., $\overline{Y}$) is the mean of the $X$ data (resp., the $Y$ data)
To compute these, we put the $X$ data in `L1` and the $Y$ data in `L2`, do `1-Var Stats L1` for $S_X$ and $\overline{X}$, similarly `1-Var Stats L2` for $S_Y$ and $\overline{Y}$, and then `LinReg(a+bx) L1,L2` for $r$. [Actually, that last one gives the complete $a$ and $b$, too.]
This yields $a = -1125.64$ and $b = 445.55$; so
$\hat{y} = -1125.64 + 445.55x$

**(b)** We want the predicted value of $Y$ – which is $\hat{y}$ – if $X$ is 4.5. Putting $x = 4.5$ into the above eqn of the LSRL gives $\hat{y} = -1125.64 + 445.55 \cdot 4.5 = 879.3361$ ($\hat{y}$ in millions of miles).

**(c)**

The form is not a big random scatter, there is a clear "U"-shape to the residual graph. Therefore the linear model is not very good!

**(d)** Following the problem instructions we make `L3`, then we compute the LSRL using the calculator command "`LinReg(a+bx) L1,L3`". This yields $\hat{y} = a + bx = 1.24593 + .27063x$.
Here is the new residual plot:

.

There is not much pattern to this residual plot, it goes up and down fairly randomly. It therefore seems that a linear model for *Log of planetary distance vs. planet number* is a fairly good fit.

**(e)** The LSRL equation in **(d)** above gives a predicted *Log of planetary distance* for the asteroid belt (using the planet number 4.5 as in **(b)**) of: $\hat{y} = a + bx = 1.24593 + .27063 \cdot 4.5 = 2.463765$.
Therefore, as the problem explains, the actual planetary distance of the asteroid belt is predicted by this model to be $10^{2.463765} = 290.914$ *million miles*.

**3** **(a)** There are really two studies here, one for *Sports Illustrated* and one for *Soap Opera Digest*. In both studies,

the OUs are "pages of the magazine"

and

the RV is binary, categorical, asking "does that page have ads?"

**(b)** For both, we use the formula

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where

$\hat{p}$ is the sample proportion (our best guess for $\pi$),
$z^*$ is the critical value for the desired confidence level.

Since the problem says to use 95% confidence level, the table gives $z^* = 1.96$.

For *Sports Illustrated*, the pages of the 9/13/1999 issue (which we are told to consider as an SRS of pages out all *SI* issues), numbered 116, out of which 54 had ads. For the 09/14/1999 *Soap Opera Digest* there were 28 pages with ads out of 130 total pages. So the sample proportions are

$\hat{p}_{si} = 54/116 = .4655$ and $\hat{p}_{sod} = 28/130 = .2154$.

So

the *SI* CI is $.4655 \pm 1.96\sqrt{.4655 \cdot .5345/116} = .4655 \pm .0908$, or $(.3747, .5563)$ [interval notation]

and

the *SOD* CI is $.2154 \pm 1.96\sqrt{.2154 \cdot .7846/130} = .2154 \pm .0707$, or $(.1447, .2861)$.

**(c)** The press might say something like

"In *Sports Illustrated*, 46% of the pages have advertisements, with a sampling error of plus or minus 9%. For *Soap Opera Digest*, $21 \pm 7$% of pages have ads."

**(d)** Yes, each interval contains the respective $\hat{p}$ – in fact, the $\hat{p}$ is the center of each interval, which has the form $\hat{p} \pm MoE$

As for the population proportion $\mu$, we have no way of knowing if it will be in any particular CI, although we expect our method is "correct" (*i.e.,* does contain $\mu$) approximately 95% of the time.

**(e)** "95% confidence" means that

If we repeatedly sample and recalculate new $\hat{p}$'s and then new CIs, using the same sample size, confidence level, and CI formulæ each time, then 95% of these new CI will contain the one, fixed (unknown) population proportion $\mu$.

**4** **(a)** The formula for CIs has a piece which is the margin of error: $MoE = z^* \sqrt{\frac{\pi(1-\pi)}{n}}$

so we need to solve $.01 = 2.576\sqrt{\frac{.15 \cdot .85}{n}}$ for $n$; here we found

the $z^* = 2.576$ in the table corresponding to a confidence level of 99%

and the problem said to use the 2005/6 figure of 15% as a starting point for the calculations for the following year. But then

$n = .15 \cdot .85 \left(\frac{2.576}{.01}\right)^2 = 8460.614$.

Actually, since it is ridiculous for $n$ not to be a whole number, we should say that

for any whole number $n \geq 8461$ the margin of error will be less than 1% with 99% confidence.

**(b)** You could either use a

**lower confidence level** – *e.g.,* with 90% confidence, $z^*$ would become 1.645 and so $n$ would be
$n \geq .15 \cdot .85 \left(\frac{1.645}{.01}\right)^2 = 3450.182$ (so $n$ any whole number $\geq 3451$)

or

**a wider target margin of error** – *e.g.,* with a goal of $MoE = .02$, we would need only
$n \geq .15 \cdot .85 \left(\frac{2.576}{.02}\right)^2 = 2115.154$ (so $n$ any whole number $\geq 2116$).

**(c)** It is important whenever you calculate a CI that your sample size $n$ and sample proportion $\hat{p}$ always obey

$$n\hat{p} \geq 10 \qquad \text{and} \qquad n(1-\hat{p}) \geq 10.$$

So after you change $n$ and resample (thus giving a new $\hat{p}$), you must recheck these conditions!