In the problems below indicate your answers by drawing boxes around them. You must clearly show your work to get credit for a problem.

1. In a random sample of 50 Texas A&M students, 27 had a pig for a pet at some time in their life. In a random sample of 40 UT students, 9 had, at some time, a pig for a pet.

   a. Estimate the true proportion of A&M students that have had pigs for pets. Give a margin of error corresponding to one standard error.

   \[ \hat{p}_1 = \text{proportion of A&M students} \]
   \[ \hat{p}_1 = 0.54 \]
   \[ \text{s.e.}(\hat{p}_1) = \sqrt{\frac{0.54 \cdot 0.46}{50}} = 0.07 \]
   \[ 0.54 \pm 0.07 \]

   b. Give a confidence interval at confidence level .9, for the difference in the percentage of A&M students that have had pig pets and the corresponding percentage of UT students.

   \[ \hat{p}_2 = \text{proportion of UT students} \]
   \[ \hat{p}_2 = 0.225 \]
   \[ \text{s.e.}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{0.54 \cdot 0.46}{50} + \frac{0.225 \cdot 0.775}{40}} = 0.097 \]
   \[ 2_{0.95} = 1.645 \]
   \[ \text{C.I. for } \hat{p}_1 - \hat{p}_2: (0.54 - 0.225) \pm 1.645 \cdot 0.097 \]
   \[ 0.315 \pm 0.16 \]
2. You want to compare the average weight of people who exercise regularly with the national average. You read that the national average weight for a man that is 5 feet 10 inches tall is 160 lbs. You take a random sample of 9 men, each about 5 feet 10 inches tall, walking out of Gregory Gym. The average weight of these men is 162 lbs with a sample standard deviation of 4 lbs. Assuming this is a random sample of 5'10" men who exercise regularly, give a confidence interval for the true average weight of such men with a confidence level of .95. Is your study evidence, at this level of confidence, that the average weight of exercising men (at that height) is different from the national average? Explain.

Small sample \( n = 9 \), heights are approximately normally distributed

Apply \( t(8) \) distribution

\[ \bar{x} = 162, \quad s = 4, \quad n = 9 \]

\[ t_{0.95} = 2.31 \quad \text{from} \quad t(8) \]

C.I. for \( \mu \) : \[ 162 \pm 2.31 \cdot \frac{4}{\sqrt{9}} \]

\[ = 162 \pm 3.08 \]

It is not evidence that the average weight is different than the national average. At this level of confidence, the true average could be any number between 158.92 and 165.08. In particular, it could be 160.
3. You hear from a friend that the proportion, \( p \), of winners at a certain slot machine is .25. To estimate \( p \), you watch as one person after another play the machine, and count that it is the 10th person that is the first to win. Find the distribution for \( p \) through a Bayesian analysis of your experiment. Then use this to find a number estimate for \( p \).

Pick an appropriate beta distribution as a prior. Note: If \( X \) is the number of times it takes to win at the slot machine in a series of plays, then \( X \) is a \( \text{Geo}(p) \) random variable. In particular, \( f_X(x) = p(1-p)^{x-1} \) for \( x = 1, 2, \ldots \)

Choose prior \( B(r,s) \) with expected value .25.

The expected value of \( B(r,s) \) is \( \frac{r}{r+s} \). So choose \( r=1, s=3 \) (other possibilities, e.g. \( r=2, s=6 \) - if you also had an estimate for the variance, this would help narrow the possibilities).

Take \( g(p) = B(1,3) \), i.e.

\[ g(p) = p^1(1-p)^2 \quad 0 < p < 1 \]

\[ L(p) = f_{\mathcal{X}}(10) = p(1-p)^9 \]

\[ h(p \mid X=10) = L(p) \cdot g(p) \]

\[ = p(1-p)^9 \cdot p^1(1-p)^2 \]

\[ = p(1-p)^{11} \]

Thus \( h(p \mid X=10) \sim \text{Beta}(2,12) \) i.e. \( f(p) \sim \frac{1}{\text{B}(2,12)} \)

We use the expected value of this distribution as a number estimate for \( p \):

\[ E(\text{B}(2,12)) = \frac{2}{2+12} = \frac{1}{7} \approx 0.14 \]
4. Let $X_1, X_2, X_3$ be a random sample from a population with distribution $f(x) = (\theta + 1)x^\theta$ for $0 < x < 1$. Let $T = \frac{X_1 + X_2 + X_3}{3}$ be an estimator for $\theta$. Find the bias for $T$.

\[
\begin{align*}
\text{bias}_T(\theta) &= E(T) - \theta \\
&= E\left(\frac{X_1 + X_2 + X_3}{3}\right) - \theta \\
&= E(X_i) - \theta \\
\text{since } X_i \text{ are identically distributed}
\end{align*}
\]

Now $E(X_i) = \int_{-\infty}^{\infty} x f_{X_i}(x) \, dx$

\[
= \int_{0}^{1} x (\theta + 1) x^\theta \, dx
\]

\[
= \frac{\theta + 1}{\theta + 2} \left. x^{\theta+1} \right|_{0}^{1}
\]

\[
= \frac{\theta + 1}{\theta + 2}
\]

Thus $\text{bias}_T(\theta) = \frac{\theta + 1}{\theta + 2} - \theta$

\[
= \frac{\theta + 1 - \theta^2 - 2\theta}{\theta + 2}
\]

\[
= \left[ \frac{1 - \theta - \theta^2}{\theta + 2} \right]
\]
5. Let $X_1, X_2, \ldots, X_{50}$ be a random sample of size 50 from a population where $X_i \in \{2, 5\}$ and $f_{X_i}(2) = 3/5; f_{X_i}(5) = 2/5$. Define $Y = X_1 + \ldots + X_{50}$. Approximate the probability that $Y$ is between 140 and 160.

By CLT

$Y \sim N(50 \mu, 50 \sigma^2)$

where

$\mu = E(\bar{X}_i), \quad \sigma^2 = Var(\bar{X}_i)$

$\mu = E(\bar{X}_i) = 2 \cdot \frac{3}{5} + 5 \cdot \frac{2}{5} = \frac{16}{5} = 3.2$

$E(\bar{X}_i^2) = 4 \cdot \frac{3}{5} + 25 \cdot \frac{2}{5} = \frac{62}{5} = 12.4$

$\sigma^2 = Var(\bar{X}_i) = E(\bar{X}_i^2) - (E(\bar{X}_i))^2 = 12.4 - 10.24 = 2.16$

Thus

$Y \sim N(160, 108)$

$P(140 \leq Y \leq 160) = P\left(\frac{-20}{\frac{108}{10}} \leq \frac{Y-160}{\frac{108}{10}} \leq \frac{0}{\frac{108}{10}}\right)$

$= P(-1.92 \leq \tilde{z} \leq 0) = .5 - .0274$

$= .4726$

6. A sample of 75 seniors from a large metropolitan area school district had a mean Math SAT score of $\bar{x} = 450$. Suppose we know that the standard deviation of the population of Math SAT scores for seniors in the district is $\sigma = 100$. A 95% confidence interval for the mean Math SAT score $\mu$ for the population of seniors is computed. Which of the following would produce a confidence interval with a smaller margin of error?

a. using a sample of 25 seniors
b. using a confidence level of 90%
c. using a sample of 50 seniors
d. using a confidence level of 99%