Problem 1 (39 pts.) A recent study published in the New England Journal of Medicine found that the amount of chocolate consumed per year per capita within a country (kg/year/capita) correlates with the number of Nobel Laureates (for every 10 million in that country’s population) with a value of $r = 0.79$.

Assume that the distribution of chocolate consumption is normal with a mean of 5.7 and a standard deviation of 3.1, and that the distribution of Nobel Laureates is also normal with a mean of 11.1 and a standard deviation of 10.

a) (10 pts.) Calculate the equation of the regression line that predicts the number of Nobel Laureates (per 10 million) as a function of chocolate consumption. Write the answer in slope-intercept form.

\[
slope = m = r \frac{s_y}{s_x} = 0.79 \frac{10}{3.1} = 2.5484
\]

\[
y' = m(x - \bar{x}) + \bar{y} = 2.5484(x - 5.7) + 11.1 = 2.5484x + 11.1 - 14.5259
\]

or $y' = 2.5484x - 3.4259$
b) (5 pts.) What is the standard error of the estimate?

\[ s_{yx} = s_y \sqrt{1 - r^2} = 10 \sqrt{1 - (0.79)^2} = 6.1311 \]

c) (5 pts.) What is the proportion of the variance in the number of Nobel Laureates (per 10 million) that can be explained by chocolate consumption?

This is the coefficient of determination:

\[ r^2 = 0.79^2 = 0.6241 \]

d) (7 pts.) Russia, which is not on the list, consumes 5.3 kg/yr per capita of chocolate. Based on the regression line, what is the expected number of Nobel Laureates in Russia (per 10 million)?

for \( x = 5.3 \), \( Y' = (2.5484)(5.3) - 3.4259 = 10.0806 \) Nobel Laureates per 10 million

e) (12 pts.) Russia actually has 1.62 Nobel Laureates (per 10 million). Assuming homoscedasticity, use the regression line and the standard error of the estimate to determine the probability that a country like Russia that consumes 5.3 kg/yr/capita of chocolate will have fewer than 1.62 Nobel Laureates (per 10 million)?

For chocolate consumption of 5.3 (kg/yr/capita), the distribution of Nobel Laureates should be normally distributed with a mean of 10.0806 and a standard deviation of 6.1311.

\[ z = \frac{y - y'}{s_{yx}} = \frac{1.62 - 10.0806}{6.1311} = -1.38 \]

\[ Pr(z < -1.38) = Pr(z > 1.38) = 0.0838 \]
Problem 2a (15 pts.) Suppose the probability that a six year old will spill his drink at the dinner table on any day is 0.45. What is the probability that he will spill his drink 5 or more times over the next 7 days?

This is a binomial problem with \( P = 0.45 \), \( n = 7 \), and \( k \geq 5 \)

Using Table B:

\[
\begin{align*}
\Pr(k=5) &= 0.1172 \\
\Pr(k=6) &= 0.0320 \\
\Pr(k=7) &= 0.0037 \\
\end{align*}
\]

\[0.1172 + 0.0320 + 0.0037 = 0.1529\]

The probability that he will spill 5 or more times in the next 7 days is 0.1529

Problem 2b (15 points) Estimate the probability that he will spill his drink 17 times or more over a period of 30 days?

Since \( n > 20 \), we will use the normal approximation to the binomial.

The number of days that he will spill will be distributed approximately normally with

mean: \( \mu = (30)(0.45) = 13.5 \)

and standard deviation: \( \sigma = \sqrt{(30)(0.45)(1 - 0.45)} = 2.7249 \)

Since we want \( \Pr(k \geq 17) \), our z-score will be

\[
z = \frac{16.5 - \mu}{\sigma} = \frac{16.5 - 13.5}{2.7249} = 1.1
\]

From Table A: \( \Pr(z > 1.1) = 0.1357 \)
Problem 3a (2 pts.) Suppose heights of sisters are correlated with \( r = 0.35 \). If you pick out a sister because she is taller than the average of the population then the other sister will be, on average,

A) as tall as the first sister  
B) less tall than the first sister  
C) taller than the first sister  
D) taller than the average of the population  
E) less tall than the average of the population  
F) [D and B]

Problem 3b (2 pts.) When is the standard error of the estimate equal to the standard deviation of \( y \)?

A) When the coefficient of non-determination is zero.  
B) When the correlation is zero.  
C) When the coefficient of non-determination is one.  
D) When the correlation is 1 or -1.  
E) D and E  
F) [B and C]

Problem 3c (2 pts.) If you see a poll that has a margin of error of \( \pm 2 \) percent, what do you know about the sample size? (assume \( P = 0.50 \))

A) [It is greater than 2000]  
B) You can’t tell the sample size.  
C) It is between 1000 and 2000  
D) It is less than 1000

Problem 4 (25 points) Suppose that you know that the age of the population of UW psychology students has a mean of 20 and a standard deviation of 2 years. Consider the mean age from a sample of 100 students.

a) (10 points) What is the mean and standard deviation of the population that this mean is drawn from?

According to the Central Limit Theorem, the sampling distribution of the mean of 100 ages will be normal with a mean of:

\[
\mu_{\bar{x}} = 20 \text{ years}
\]

and a standard deviation of

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{100}} = 0.2 \text{ years}.
\]

b) (15 points) What is the probability that the mean age from a sample of 100 students will exceed 20.3 years?

The z-score for a mean of 20.3 years is

\[
\frac{x - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{20.3 - 20}{0.2} = 1.5
\]

From table A, \( Pr(z > 1.5) = 0.0668 \)