Homework Assignment 3
Solutions

1. Please see the code. The 95% confidence interval for the population mean is (92.1;109.5), the 95% confidence interval for the population standard deviation is (5.2;20.3), and the 95% confidence interval for the population variance is (26.7;412.5).

[ 4 points ]

2. The sample mean is 103 and the sample SD is 10.67 for strain A, the sample mean is 67 and the sample SD is 9.30 for strain B. Assuming the within strain population standard deviations are equal, we derive a pooled estimate of the population SD as \( \sqrt{(10.67^2 \times 9 + 9.30^2 \times 4)/13} = 10.27 \). The estimated SE of the difference between the sample means is \( 10.27 \times \sqrt{1/10 + 1/5} = 5.62 \). The 97.5 percentile of the t distribution with 13 degrees of freedom is 2.16, the 99.5 percentile is 3.01. The 95% confidence interval is \( (103 - 67) \pm 2.16 \times 5.62 = 36 \pm 12.1 = (23.9, 48.1) \). The 99% confidence interval is \( (103 - 67) \pm 3.01 \times 5.62 = 36 \pm 16.9 = (19.1, 52.9) \). The p-value is \( 2 \times 10^{-5} \). Also see the code.

[ 4 points ]

3. (a) We have \( n = 400, \bar{x} = 140, s = 25 \). Since \( n \) is very large, we can use a normal approximation, and the 95% confidence interval can be approximated by \( 140 \pm 1.96 \times 25/\sqrt{400} = 140 \pm 2.44 = (137.56; 142.44) \). The 99% confidence interval can be approximated by \( 140 \pm 2.58 \times 25/\sqrt{400} = 140 \pm 3.23 = (136.77; 143.23) \).

(b) We reject the null hypothesis if the test statistic is less than -1.96, or larger than 1.96. With the observed data, the test statistic is \( (\bar{x} - 130)/(s/\sqrt{n}) = 10/1.25 = 8 \). Thus, we reject the null hypothesis.

[ 4 points ]

4. Please see the code. The 95% confidence interval for the treatment effect is (46.8;152.8). The p-value is 0.005. We conclude that there is a treatment effect.

[ 2 points ]

5. (a) The expected mean is

\[ E(\hat{p}) = p = 0.3, \]

and the variance is

\[ \text{Var}(\hat{p}) = \frac{p \times (1 - p)}{n} = \frac{0.3 \times (1 - 0.3)}{100} = 0.0021. \]

(b) See the code.

(c) I get a sample mean of 0.300045, and a sample variance of 0.002101, so very close to the theoretical values above.
See the code. The distribution of the proportions observed in our simulation seem to be close to a Normal distribution. A good rule of thumb is that if \( n \times p \times (1 - p) > 5 \), the normal approximation holds. Here, \( n \times p \times (1 - p) = 21 \).

6. Please also see the code.

(a) For any single person we have

\[
P(X < 80) = P((X - 90)/5 < (80 - 90)/5) = P(Z < -2) = 0.0228,
\]

where \( X \sim N(\mu = 90, \sigma = 5) \) and \( Z \sim N(0,1) \). Thus, among 1,000 people randomly selected from this population, the expected number of people with FPG less than 80 mg/dl is \( n \times p = 1000 \times 0.0228 = 22.8 \), based on a Binomial distribution with \( n = 1000 \) and \( p = 0.0228 \).

(b) For 25 people we have \( \mathbb{E}[\bar{X}] = 90 \) and \( \text{Var}[\bar{X}] = \text{Var}[X]/25 = 1 \). Therefore

\[
P(\bar{X} > 92) = P((\bar{X} - 90)/1 > (92 - 90)/1) = P(Z > 2) = 0.0228.
\]

(c) For any single person we have

\[
P(X > 100) = P((X - 90)/5 > (100 - 90)/5) = P(Z > 2) = 0.0228.
\]

Let \( Y \) be a Binomial distribution with \( n = 5 \) and \( p = 0.0228 \). Then

\[
P(Y \geq 4) = P(Y = 4) + P(Y = 5) = \binom{5}{4} \times 0.0228^4 \times (1 - 0.0228) + 0.0228^5 = 1.3 \times 10^{-6}.
\]