Homework #4

1. (10 points) A sample of size 100, which has the sample mean $\bar{X} = 500$, was drawn from a population with an unknown mean μ and the standard deviation $\sigma = 80$.

a) What is the probability that the population mean will be in the interval (480, 510)?

Answer: $Z = \dfrac{\mu - 500}{80/\sqrt{100}}$ follows normal distribution by approximation. Hence,

$$P(480 < \mu < 510) = P\left(\frac{480-500}{80/\sqrt{100}} < Z < \frac{510-500}{80/\sqrt{100}}\right) = P(-2.5 < Z < 1.25) = 0.8881$$

b) Give the 95% confidence interval for the population mean.

Answer: $P(500 - z_{0.025}\dfrac{80}{\sqrt{100}} < Z < 500 + z_{0.025}\dfrac{80}{\sqrt{100}}) = 0.95$ where $z_{0.025} = 1.96$.

Therefore, the interval is [484.32, 515.68].

2. (10 points) Find the maximum likelihood estimate for λ in a sample of size n drawn the Poisson distribution

$$f(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Answer: The log likelihood function is $L(\lambda) = \sum_{i=1}^{n}(x_i \ln\lambda - \lambda - \ln x_i!)$. We need to find the maximum by finding the derivative $L'(\lambda) = \dfrac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0$ which implies that the estimate should be $\dot{\lambda} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

3. (18 points) Among all the computer chips produced by a certain factory, 6% are defective. A sample of 400 chips is selected for inspection.

a) Use Matlab or other computational software to find the exact probability that this sample contains between 20 and 25 defective chips (including 20 and 25)? Hint: the number of defective chips follow binomial distribution.

Answer: In Matlab, type in

pdf('binomial',20,400,0.06)+pdf('binomial',21,400,0.06)+pdf('binomial',22,400,0.06)+pdf('binomial',23,400,0.06)+pdf('binomial',24,400,0.06)+pdf('binomial',25,400,0.06). The value Matlab returns is 0.46.

b) Use the central limit theorem to approximate the same probability. Hint: to account for rounding error integrate the appropriate normal distribution between 19.5 and 25.5.

Answer: Let X be the number of defective chips in the sample of 400. Then $X \sim \text{Binomial}(n = 400, p = 0.06)$. Approximately,

$X \sim \text{Normal}(\mu = 400,{*}0.06 = 24, \sigma = \sqrt{400 * 0.06 * (1-0.06)} = 4.75)$.

$P(19.5 < X < 25.5) = P(-0.95 < Z < 0.32) = 0.45$

c) Suppose that 40 inspectors independently from each other collected samples of 400 chips each. What is the probability that at least 14 inspectors will find between 20 and 25 defective chips in their samples? Hint: use normal distribution CDF at 13.5 (instead of 14) to approximate the binomial distribution.

Answer: Let Y be the number of inspectors who will find between 20 and 25 defective chips in their samples. $Y \sim \text{Binomial}(n = 40, p = 0.45)$. Approximately

$Y \sim \text{Normal}\ (\mu = 40 * 0.45 = 18, \sigma = \sqrt{40 * 0.45 * (1-0.45)} = 3.15)$

$P(Y > 13.5) = P(Z > (13.5-18)/3.15 = -1.43) = 1 - P(Z < 1.43) = 0.92$

4. (10 points) (10 points) The elasticity of a polymer is affected by the concentration of a reactant. When low concentration is used, the true mean elasticity is 55, and when high concentration is used the mean elasticity is 60. The standard deviation of elasticity is 4 in the first case and 5 in the second case. Random samples of sizes 16 and 9 correspondingly are taken, find the probability that $\bar{X}_{high} - \bar{X}_{low} \geq 4$.

Answer: Let $\Delta x = \bar{X}_{high} - \bar{X}_{low}$ denote the sample difference, which approximately follows normal distribution with mean equal to 60-55=5 and standard deviation $\sigma = \sqrt{\frac{4^2}{16} + \frac{5^2}{9}} = 1.94$. Using z-table, we could find $P(\Delta x \geqslant 4) = P\left(z = \frac{\Delta x - 5}{1.94} \geqslant -0.52\right) = 0.70$.

5. (12 points) To estimate the copy number of a specific protein, a laboratory has done multiple measurements:

2310, 2320, 2010, 10800, 2190, 3360, 5640, 2540, 3360, 11800, 2010, 3430, 10600, 7370, 2160, 3200, 2020, 2850, 3500, 10200, 8550, 9500, 2260, 7730, 2250

(a) Find a point estimate of the mean protein copy number.

Answer: $\dot{\mu} = \frac{\sum_{i=1}^{25} x_i}{25} = 4958.4$

b) Find a point estimate of the standard deviation of the protein copy number

Answer: $\dot{\sigma} = \sqrt{\frac{\sum_{i=1}^{25}(x_i - \bar{X})^2}{25-1}} = 3420.5$

c) What is approximately the standard error of the estimate of the mean protein copy number obtained in part a)

Answer: $SE_{\bar{\mu}} = \frac{\dot{\sigma}}{\sqrt{n}} = 684.2$

d) Find a point estimate for the proportion of readings that are less than 5000.

Answer: $\text{Estimate} = \frac{\text{\# of readings smaller than 5000}}{25} = \frac{16}{25}$

e) Find 95% confidence intervals for the point estimate in part d)

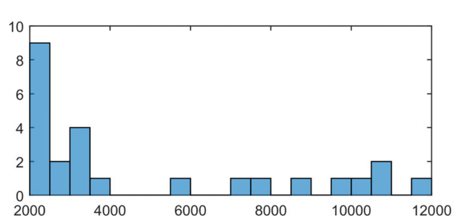
Answer: The proportion of readings follows the Bernoulli distribution with p=0.64. Using normal approximation, the z-variable is constructed by

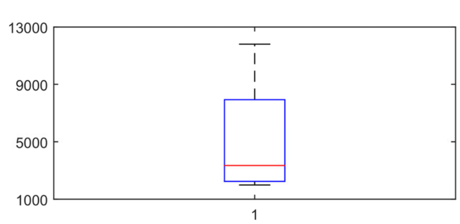
$$z = \frac{\mu - 0.64}{\sqrt{0.64 * (1 - 0.64)/25}} = \frac{\mu - 0.64}{0.096}$$

. Hence, $P(0.64 - z_{0.025}0.096 < z < 0.64 + z_{0.025}0.096) = 0.95$ where $z_{0.025} = 1.96$ and the interval is thus given by [0.45, 0.83].

Note: the standard error of the sample mean is $\sqrt{p(1-p)/n}$.

f) Use the computer to plot the histogram and the box-and-whisker for the sample.





figure; hist(sample, n_of_bins);

figure; boxplot(sample);