Homework 5 Solution

This assignment is worth 35 points.

1. Working for a car company, you have been assigned to find the average miles per gallon (mpg) for a certain model of car. You take a random sample of 15 cars of the assigned model. Based on previous evidence and a QQ plot, you have reason to believe that the gas milage is normally distributed. You find that the sample average miles per gallon is around 26.7 with a standard deviation of 6.2 mpg.

(a) Construct and interpret a 95% confidence interval for the mean mpg, μ, for the certain model of car.

(4pts)Solution Keeping in mind that with a sample size of 15 and a normal distribution we would want to use t distribution to construct the CI, we solve: $(26.7 \pm 2.145\frac{6.2}{\sqrt{15}})$ mpg $= (23.266, 30.134)$ mpg. If we had taken many samples of size 15 from this population, and created a confidence interval by this method for each sample, 95% of them would cover the true average mpg for that particular model of car.

(b) What would happen to the interval if you increased the confidence level from 95% to 99%? Explain your reasoning. (Do not compute the new interval.)

(3pts) The CI would grow wider. This makes sense because confidence-level and width of CI have an inverse relationship (with n constant). To be more sure that our estimate is in fact in the interval, we would need to grow that interval to account for random error our estimate.

(c) The lead engineer is not happy with the interval you contructed and would like to keep the width of the whole interval to be less than 4 mpg wide. How many cars would you have to sample to create the interval the engineer is requesting?

(4pts) We solve: $\frac{2^2(1.96)^2(6.2)^2}{4^2} = 36.92$, i.e. we need **at least** a sample of 37 cars.

(d) Suppose you are asked to repeat this process with data gathered from all across the country. You end up constructing a total of 60 separate 95% confidence intervals at different factories for the estimated average mpg of the model of car. Of these intervals, how many of them do you expect would fail to contain the true value of μ?

(2pt) $(0.05) * (60) = 3$ intervals.

2. Suppose a computer engineer is interested in determining the average weight of a motherboard manufactured by a certain company. A summary of a large sample provided to the engineer suggest a mean weight of 11.8 ounces and an estimated standard deviation of 0.75 ounces.

(a) How large a sample size is required if we want a 99% confidence interval, with a tolerable interval width of 0.4?

(4pts) We solve: $\frac{2^2(2.58)^2(.75)^2}{.4^2} = 93.61$, i.e. a sample of size 94 is required.

(b) How large a sample would we need if we are interested in a 95% confidence interval with a tolerable width of 0.5?

(4pts) $\frac{2^2(1.96)^2(.75)^2}{.5^2} = 34.57$, a sample of size 35 is required.

3. A large population has an unknown distribution, but it is known that the population has mean 100 and variance 200. A sample of size n = 90 is going to be taken. Use the CLT to find the approximate probability that the sample mean will be greater than 102.

(4pts) The sample size is quite large, so we will apply Central Limit Theorem (CLT). The distribution of the sample mean is thus $\bar{X} \tilde{\sim} N(100, 200/90)$. Thus $P(\bar{X} > 102) = P(\frac{\bar{X}-100}{\sqrt{200/90)}} > \frac{102-100}{\sqrt{200/90}}) = P(Z > 1.34) = 1 - 0.9099 = 0.0901$.
Alternatively, find $P(\bar{X} > 102)$ from R:

```
library(distributions3)
```

```
##
## Attaching package: 'distributions3'

## The following objects are masked from 'package:stats':
##
##     Gamma, quantile

## The following object is masked from 'package:grDevices':
##
##     pdf
```

```
Xbar <- Normal(100, sqrt(200/90))
1 - cdf(Xbar, 102)
```

```
## [1] 0.08985625
```

(Note: the discrepancy is due to rounding of $\frac{102-100}{\sqrt{200/90}} \approx 1.34$.)

4. Most penguin species are not sexually dimorphic, which means that there are no obvious outward body characteristics which indicate gender. Therefore, penguin gender must be determined by a blood test. A penguin researcher is interested in estimating the proportion of females in a large penguin population. She takes a random sample of n = 24 penguins, and determines the gender of each one using a blood test. She finds 15 males and 9 females. Let π be the proportion of females in the population.

(a) Compute a numerical point estimate of π.

(3pts) Suppose X denotes no. of female penguins in a random sample of size 24. Then $X \sim Bin(24, \pi)$. Hence $E(X) = 24\pi$ where the realized value of X is 9. Therefore a numerical point estimate of π is the sample proportion of females which is $p = 9/24 = 0.375$.

(b) Compute the estimated standard error of the estimate.

(4pts) The estimated standard error is $\sqrt{\frac{(0.375*0.625)}{24}} = 0.099$.

(c) Is it possible to compute a 95% CI for π using the normal approximation in this case? If it is possible, explain why, and make the CI. If it is not possible, explain why not.

(3pts) Since $24(0.375) = 9 > 5$ and $24(0.625) = 15 > 5$, we can use the CLT to approximate the distribution of $\hat{\pi}$ as a normal. Thus the CI is $0.375 \pm 1.96(0.099)$ or $(0.181, 0.569)$.