ST 516 - Homework 4
student Paul ReFalo 10/19/17

1. The column motheriq contains the mother's IQ for 36 gifted children. We are interested in whether the mothers of gifted children have an IQ higher than the population at large, which is 100.

   . (a) State the null and alternative hypothesis in statistical notation, and in words.

$$H_0 : \mu = 100$$

$$H_A : \mu > 100$$

The null hypothesis is that the mothers of gifted children have average IQ's equal to that of the population average of 100.

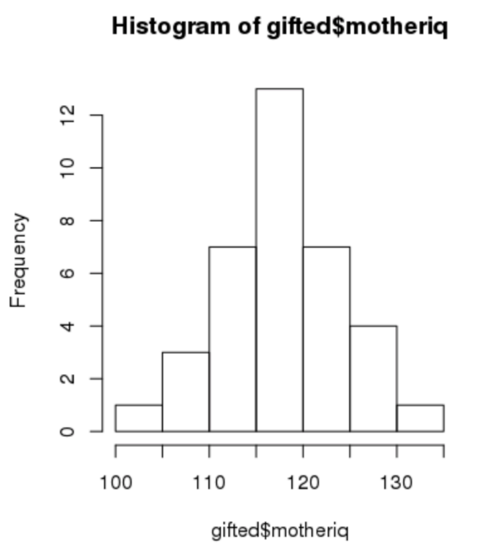
The alternative hypothesis is that the mothers of gifted children have average IQ's greater than the population average of 100.

Continued on next page

.   (b) Give the formula for the test statistic you will use, and calculate it.

I will use the sample mean to estimate the population mean. The formula for the sample mean is:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$



We can use the sample mean since the motheriq data has no outliers and shows no indication of strong skew and has n = 36.

```
> hist(gifted$motheriq)

> mean(gifted$motheriq)

[1] 118.1667
```

The sample mean is an estimate for the population mean and is unbiased.

Continued on next page

(c) Give the p-value for the test, and the line of code you used to calculate it.

See part d for the calculate of the Z value

```
> pval <- 1 - pnorm(16.82) # Z value, one-sided or upper tail
> pval # practically I would report this as pval < 0.000001
[1] 0
```

. (d) Calculate a point estimate and a 95% confidence interval for the mean IQ of mothers of gifted children.

Our point estimate, $\bar{x} = 118.1667$

$$SE_{\bar{x}} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

point estimate $\pm$ 1.96×*SE*

```
> sd(gifted$motheriq)
[1] 6.504943
```

SE = 6.504943 / sqrt(36)

SE = 1.084157

point estimate ± 1.96 x 1.084157 at 95% Confidence Interval

118.1667 ± 2.12 → (116.05, 120.29) at 95% Confidence Interval

$Z = \bar{x} - \mu_0 / (s2 / n)^{1/2} = 118.17 - 100 / (6.5^2 / 36)^{1/2} = 18.17 / 1.08$

$Z = 16.82$ and since $|Z| > 1.96$ we reject H_0

(e) Give a summary of your findings.

We are 95% confident that the population mean of mothers of gifted children is between an IQ of 116.05 and 120.29. Because the average IQ of 100 is not inside of this range we reject the null hypothesis in favor of the alternative hypothesis. Since the 95% confidence interval is well above 100 we can conclude that the average, or mean, IQ of mothers of gifted children is greater than 100.

2. (2 points) In your own words, describe how understanding the behavior of a statistic over many samples (the sampling distribution) allows us to make inference about a population from a single sample.

The goal of statistical inference is to be able to make conclusions about a population based upon a sampling of that population.

From Lecture 1

> Statistical inference is the process of learning—or inferring—properties of a population distribution from properties of a sample from that distribution.

The Central Limit Theorem tells us that so long as the sample size is sufficient, typically about 30 or greater for populations aren't strongly skewed, we know that the mean of samples of a population will be normally distributed and equal to that of the larger population. If the variance is large and the distribution of the sample shows skew then a larger sample size may be needed. As sample size increases we expect the sample distribution to narrow or decrease. Having this normal distribution of a test statistic is the basis for hypothesis testing thereby giving a method to make statistical inferences about the population.

Continued on next page

4. (1 point) A random sample of 10 college graduates is selected to play Sudoku, with each graduate attempting the same puzzle. The mean time it takes them to correctly complete the puzzle is 5 min.

   . (a) Which number should we use to estimate the population mean time it takes college graduates to finish this puzzle? Why?

   The sample mean because it is an estimate for the population mean and is unbiased.

   . (b) Why is there uncertainty in this estimate?

   There is uncertainty because this is a sample and not a survey of the entire population of college graduates, which would be impractical to obtain.

   . (c) What happens to the uncertainty if we increase the sample size to 1000 graduates?

   The uncertainty will decrease as the sample size increases because the sample distribution will narrow.

5. (1 point) A farmer in Idaho estimates that the mean time it takes to grow one potato is 100 days. His 95% confidence interval is (90 days, 110 days).

   . (a) What is wrong with the following interpretation of this confidence interval? There is a 95% chance that the mean time it takes to grow one potato is between 90 and 110 days.

   It is inappropriate to apply the confidence interval to a given data point or one potato.

   . (b) Describe one correct way to interpret this confidence interval.

   It is appropriate to say to a 95% confident that the mean time to grow a potato is between an 90 and 110 days.

   . (c) Describe what this confidence interval means in terms of probability.

   The confidence interval means the probability of failing to reject the null hypothesis and not making a Type I error.

. (d) How would the confidence interval change if the confidence level changes?

As the confidence level increases we are casting a wider net so the confidence interval gets wider.

6. (1 point) A local restaurant is worried its potato supplier is skimping on the bags of potatoes it supplies. They claim each bag weighs 15lbs. The restaurateur understands that its hard to get a whole number of potatoes to weigh exactly 15lbs, but they feel it should average out. Over a month they weigh every bag they receive (and assume this is a random sample from all bags). They find a 95% confidence interval of the mean weight is 14.5 lbs to 14.9 lbs.

. (a) Would they reject or fail to reject the null hypothesis (at the 5% level) that the mean weight is 15lbs?

Reject the null hypothesis (bags' mean weight = 15 lbs) in favor of the alternative hypothesis (bags' mean weight ≠ 15 lbs).

. (b) What can you say about the p-value from the hypothesis test?

The p-value = 1.96 * Z ≤ 0.05 for 95% confidence interval.

. (c) What can you say about the sample mean from the measured bags?

The sample mean should be normally distributed about the midpoint of the confidence interval or 14.7 lbs.

. (d) What can you say about the p-value from the hypothesis test where the null hypothesis is the mean weight is 14.8 lbs?

The p-value = 1.96 * Z > 0.05 for 95% confidence interval.

. (e) The restaurateur presents this evidence to the supplier and asks for a refund. If you were the supplier, are there any problems you would bring up that cast doubt on the analysis?

At a higher confidence interval like 99% the result of the analysis may then be to fail to reject the null hypothesis that the bags' weight = 15 lbs. Also, the bags sampled may not represent a true random sampling of the bags from the supplier. Also a larger sample size should narrow the sample distribution around the true population mean and change the analysis.