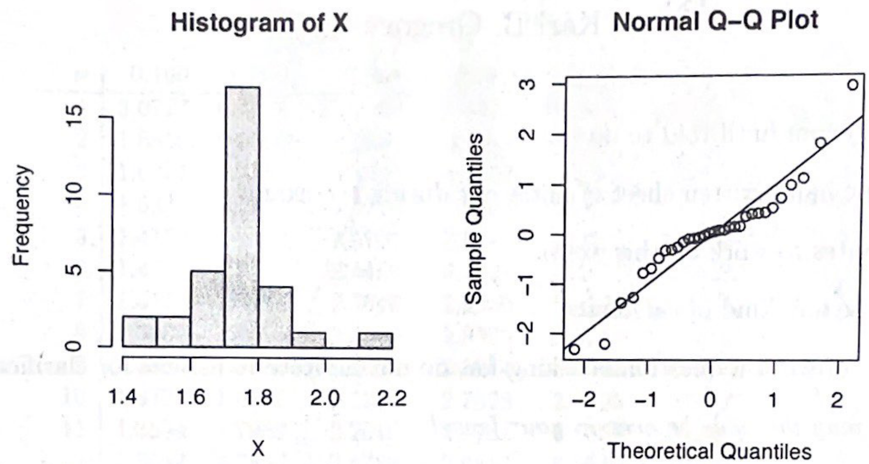
1. The length and diameter of each acorn in a sample of 32 live oak acorns was measured (in this class!!). Suppose the ratio of the length to the diameter is of interest. The 32 length-to-diameter ratios in the sample had mean $\bar{X}_n = 1.74$ and standard deviation $S_n = 0.138$. The figure below shows a histogram and a normal QQ plot of the length-to-diameter ratios.



(a) Explain the purpose of the normal QQ plot and give your interpretation of this one.

The purpose is to decide if we can assume Normality of the sampling distribution. It does appear to be mostly Normal as the plots don't deviate very heavily. However, even if this plot was determined to be non-Normal we have a sample size ≥ 30 so we can apply the CLT regardless.

(b) A 95% confidence interval for the mean length-to-diameter ratio of live oak acorns is constructed with an formula like this:

$$(i) \pm (ii)\frac{(iii)}{(iv)}.$$

I assumed normality. therefore I will use $t_{n-1,\alpha/2}$ for (ii)

Give the numbers to plug in for (i), (ii), (iii), and (iv).

i) 1.74 ii) 2.0395 iii) 0.138 iv) $\sqrt{32}$

$$1.74 \pm 2.0395\frac{0.138}{\sqrt{32}}$$

[redacted]

(c) Of the three intervals $(1.69, 1.79)$, $(1.70, 1.78)$, and $(1.67, 1.81)$, one is the 90% confidence interval, one is the 95% confidence interval, and one is the 99% confidence interval based on these data for the mean length-to-diameter ratio of live oak acorns. Which interval is the 90% confidence interval?

The 90% CI interval will be the most narrow of the three which means it is the (1.70, 1.78) CI

(d) Suppose one wished to test whether the mean length-to-diameter ratio of live oak acorns was the golden ratio 1.618. Give the hypotheses of interest, using μ to denote the mean length-to-diameter ratio of the live oak acorn population.

$H_0: \mu = 1.618$

$H_1: \mu \neq 1.618$

(e) The test statistic for testing the hypothesis is computed with a formula like this:

$$T_{\text{test}} = \frac{(i) - (ii)}{\sqrt{(iii)/(iv)}}.$$

Give the numbers to plug in for (i), (ii), (iii), and (iv).

$(i) = 1.74$

$(ii) = 1.618$

$(iii) = 0.138^2$

$(iv) = 32$

(f) The test statistic value is $T_{\text{test}} = 5.011806$. Give your conclusion about the golden ratio hypothesis using significance level $\alpha = 0.01$.

Reject H_0 if $|T_{test}| > t_{n-1,\alpha/2}$

We REJECT H_0 at the significance level $\alpha = 0.01$

$T_{Test} = 5.011806$

$t_{32-1,.005} = 2.7440$

$\Rightarrow 5.011806 > 2.7440$

$$\begin{array}{r} 38 \\ -\ 11 \\ \hline 27 \end{array}$$

2. In a survey of 38 students (in this class!!), 11 reported that they had a houseplant. Let's regard the 38 students as a random sample of USC students.

(a) The Wald-type 95% confidence interval for the proportion of USC students with a houseplant is constructed with a formula like this:

$$(i) \pm (ii)\sqrt{\frac{(iii)}{(iv)}}.$$

Give the numbers to plug in for (i), (ii), (iii), and (iv).

$i = \hat{p}_n = 11/38$

$ii = z_{\alpha/2} = 1.96$

$iii = \hat{p}_n(1-\hat{p}_n) = 11/38(38/38 - 11/38) = 11/38(27/38)$

$iv = n = 38$

$$11/38 \pm 1.96\sqrt{\frac{11/38(27/38)}{38}}$$

(b) For the Agresti-Coull interval (which has much better performance), we add two "successes" and two "failures" to the data set and recompute the Wald-type interval. Give the numbers (i), (ii), (iii), and (iv) such that

$\tilde{p}_n = \frac{11+2}{38+4} = \frac{13}{42}$

$$(i) \pm (ii)\sqrt{\frac{(iii)}{(iv)}}$$

gives the Agresti-Coull interval for the proportion of USC students with a houseplant.

$i = \tilde{p}_n = 13/42$

$ii = 1.96$

$iii = \tilde{p}_n(1-\tilde{p}_n) = 13/42(1 - 13/42)$

$iv = n+4 = 38+4 = 42$

$$13/42 \pm 1.96\sqrt{\frac{13/42(1-13/42)}{42}}$$

(c) The 95% Agresti-Coull interval is (0.170, 0.450). Give an interpretation of this interval.

~~We can say~~

We are 95% sure that we have captured the true proportion of USC students w/ housplants in the interval (0.17, 0.45)

(d) Suppose you wish to more accurately estimate the proportion of USC students with houseplants. Specifically, suppose you wish to estimate it within 1 percentage point with 99% confidence. Give an expression for the sample size required (you do not have to simplify your expression).

$$n = \left(z_{\alpha/2} \cdot \frac{\sqrt{p(1-p)}}{M} \right)^2$$

$$n = \left(2.576 \cdot \frac{\sqrt{\left(\frac{11}{38}\right)\cdot\left(\frac{27}{38}\right)}}{0.01} \right)^2$$

(e) The required sample size from part (d), if we use the survey data to make a guess at the population proportion, comes out to $n = 13{,}647$, which you decide is too large. How can you change your specifications in part (d) to make the required sample size smaller?

Increase Margin of Error > 0.01

Decrease Confidence Level $< 99\%$

(f) Suppose your botany professor claims that no more than 10% of USC students have houseplants. Give the null and alternate hypotheses for testing his claim.

$H_0 \leq 0.1$

$H_1 > 0.1$

(g) For testing the hypotheses in part (f), suppose you compute

$$Z_{\text{test}} = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}} = 2.407$$

using the survey data. Give the corresponding p-value.

p-value = 0.008 $\quad z > 2.41 \quad 0.5 - z(2.41)$

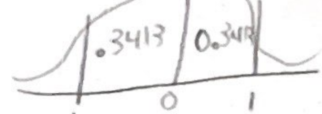
(h) Give your conclusion about the claim of the botany professor in part (f). Use significance level $\alpha = 0.05$.

p-value less than α

∴ Reject the claim of professor.

Reject H_0

$\bar{X}_n = 135\,g \quad \sigma = 15\,g$



3. Suppose the weights of bananas on sale at your grocery store have a Normal distribution with a mean of 135 grams and a standard deviation of 15 grams.

   (a) Give the probability that a randomly selected banana weighs between 120 and 150 grams.

$Pr(120 < X < 150) = \quad Z = \frac{\bar{X}_n - X}{\sigma} \Rightarrow Z = \frac{135-120}{15} \Rightarrow Z = 1 = 0.3413 \times 2 = 0.6826$

$Z = \frac{135-150}{15} \Rightarrow Z = -1 \quad (-1 < Z < 1) \quad \boxed{Pr(120 < X < 150) = 0.6826}$

   (b) Give the probability that the mean of the weights of 9 randomly selected bananas falls between 120 and 150.

$Pr(120 < \mu < 150) = \quad Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \Rightarrow Z = \frac{135-120}{15/\sqrt{9}} = \frac{15}{5} = 3 \quad \boxed{Pr(120 < \mu < 150) = 0.9874}$

$Z = \frac{135-150}{15/\sqrt{9}} = -3 \quad (-3 < Z < 3) \quad Z = 3 = 0.4987$

$\times 2 = 0.9874$

   (c) Give an explanation for why there is a difference between the answers to parts (a) and (b).

There is a difference because in the first part we are dealing with one randomly selected banana, so the probability distribution is different, whereas in part (b) we have 9 bananas and are much more likely to have a mean fall in that range.

4. Students taking a survey (in this class!!) were asked to weigh their keychains and record the weight in grams. Thirty-five students weighed their keychains. The mean weight was 84.71 grams. Consider the three sets of hypotheses: $n = 35 \quad \bar{X}_n = 84.71$

(1)	(2)	(3)
H_0: $\mu \geq 70$	H_0: $\mu = 70$	H_0: $\mu \leq 70$
H_1: $\mu < 70$	H_1: $\mu \neq 70$	H_1: $\mu > 70$

When the survey data are used to test these sets of hypotheses, the tests result in the p-values below; match each p-value to one of the hypotheses (1), (2), or (3).

   (a) The p-value 0.0434. hypothesis (3)

   (b) The p-value 0.9566. Supports the null, hypothesis (1)

   (c) The p-value 0.0868. hypothesis (2)

Since it's just double the p-value from (3).