7 More One-Sample Confidence Intervals and Tests, Part 1 of 2

We already have a Z confidence interval (§5) and a Z test (§6) for an unknown mean μ for when we know σ and have a normal population or large n. In this unit we study:

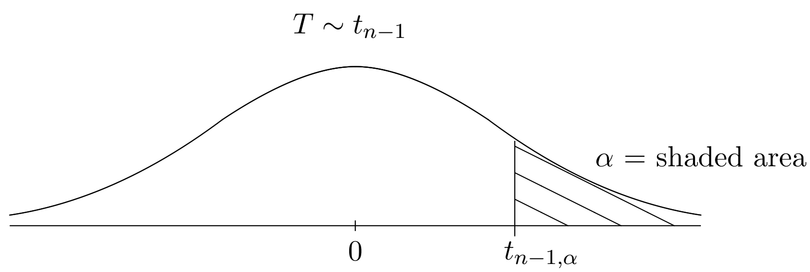
- the *Student's* t_{n-1} distribution of $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$, useful with ____________ (compare T to $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$); and a T CI and test for μ for a normal population or large n and an unknown σ
- the relationship between a two-sided confidence interval and a ____________________
- *power* $= P(\text{reject } H_0|$____________________$)$
- a *bootstrap* CI and test for μ requiring only a simple random sample
- a *sign* test for an unknown _____________ M requiring only a simple random sample
- a Z CI and test for an unknown proportion π
- extra examples (if time allows)

The Student's t Distribution

Now suppose we have a normal $\bar{X} \sim N(\mu, \sigma^2/n)$ are interested in μ but don't know σ. Define the random variable $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$. T's distribution isn't normal; it's the *Student's t distribution with $n-1$ degrees of freedom*, denoted t_{n-1}. ("Student" is a pseudonym for William Gosset, a statistician at ____________.)

- Properties of $T \sim t_{n-1}$:
  - T is a sample version of a ____________, estimating how far $\bar{X}$ is from __________________, in _________________________
  - t_{n-1} looks like $N(0,1)$: symmetric about ____, ____________-peaked, and ________-shaped
  - T's variance is ____________ than Z's because estimating σ (_________) by S (___________) gives T more variation than Z: t_{n-1} is shorter with thicker tails (draw $N(0,1)$ and t_{6-1})
  - As n increases, t_{n-1} gets closer to ____________ (S becomes a __________________ of σ); in the limit as $n \to \infty$, they're ____________
- t table:

  Let $t_{n-1,\alpha}$ = the critical value t cutting off a __________________ area of α from t_{n-1} (draw). The posted Student's t table gives ____________ tail probabilities.



e.g. Use the t table to find the critical value t

- cutting off a right tail area of .05 from the t_{6-1} distribution: $t_{5,.05} =$ ____________
- such that the area under the t_{22-1} curve between $-t$ and t is 98%
- such that the area under the t_{25-1} curve left of t is .025

Confidence interval using the Student's t distribution

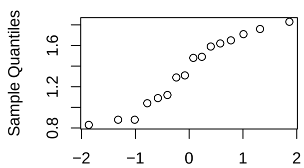
Theorem:

If $X_1, \ldots, X_n$ is a simple random sample, from $N(\mu, \sigma^2)$ or where n is large (say $n > 30$), then $\bar{X} \pm t_{n-1,\alpha/2} \frac{S}{\sqrt{n}}$ contains μ for a proportion $1 - \alpha$ of random samples.

Proof:

e.g. Recall §5 example with a warehouse of thousands of painted engine blocks; a random 16 have paint thickness measured, with $\bar{x} = 1.348$ and $s = .339$. Make a 95% confidence interval for μ, the unknown population mean thickness (supposing now we don't know σ).

$n =$ ______ ; Is n large enough or is sample from normal population? Try `qqnorm(paint)`:



We have $1-\alpha =$ ____ $\implies \alpha =$ ____ $\implies t_{n-1,\alpha/2} =$ ______________,

so $\bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} =$

With what probability does our interval contain μ? _____________

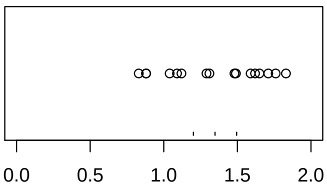
e.g. In a sample of 100 boxes of a certain type, the average compressive strength was 6230 N, and the standard deviation was 221 N.

a. Find a 95% confidence interval for the mean compressive strength.

b. Find a 99% confidence interval for the mean compressive strength.

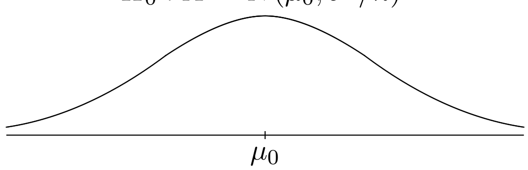
Hypothesis test using the Student's t distribution

Now suppose an engine design specification says the paint thickness should be 1.50 mil. We want to know whether the device is off this mark on average, so that it should be re-calibrated to correct its population mean thickness, μ. We test H_0 : _______ vs. H_A : _______.



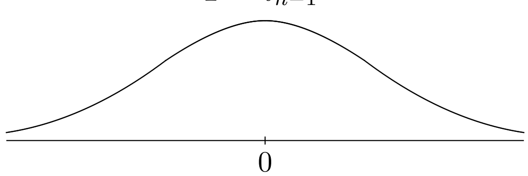
If the population is normal or n is large enough that the CLT applies, and if H_0 is true, then $\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})(\approx)$, where μ_0 is the value of μ under _______

$H_0 : \bar{X} \sim N(\mu_0, \sigma^2/n)$



That means $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$:

$T \sim t_{n-1}$



Values of $\bar{X}$ far from μ_0 (in ______________________________), or equivalently, values of t far from _______ indicate strong evidence against H_0.

Let's use significance level $\alpha = .05$. We have three options for completing the test.

- Find the rejection region corresponding to a chosen significance level $\alpha = P(\text{type I error}) =$ ____. This region is $T < -t_{n-1,\alpha/2}$ or $T > t_{n-1,\alpha/2}$ (draw). Compute t and reject H_0 if it is in the region.

  For the paint, we have $n = 16$, so we need $t_{16-1,0.025} =$ ______________. Our rejection region is ____. Our observed t is
  $t_{obs} = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} =$ ______________
  Conclusion:

- Compute the p-value and compare it to α.

  We have $t_{obs} = -1.796$, so our p-value is (draw)

  p-value $= P(T \text{ is as extreme or more extreme than } t_{obs} | H_0 \text{ is true}) =$ ______________

  - from the t table, ______________, or
  - from R, use `2 * pt(q = -1.796, df=15)` = ______________

  Conclusion:

- Compute a confidence interval and check whether μ_0 is in it (below).

Summary:

Suppose $X_1, \cdots, X_n$ is a simple random sample from $N(\mu, \sigma^2)$ or n is large.

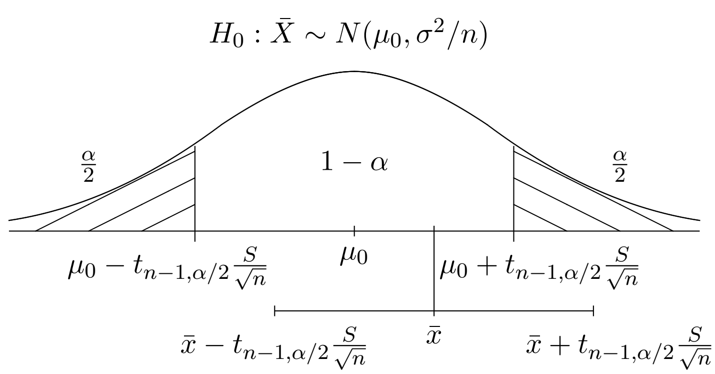
- To test that μ has a specified value, $H_0 : \mu = \mu_0$,
  1. State null and alternative hypotheses, H_0 and H_A
  2. Check assumptions
  3. Find the test statistic, $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$
  4. Find the p-value, which is an area under t_{n-1} depending on H_A:
     $H_A : \mu > \mu_0 \implies p\text{-value} = P(T > t)$, the area right of t (where $T \sim t_{n-1}$)
     $H_A : \mu < \mu_0 \implies p\text{-value} = P(T < t)$, the area left of t
     $H_A : \mu \neq \mu_0 \implies p\text{-value} = P(T < -|t|) + P(T > |t|)$, the sum of areas left of $-|t|$ and right of $|t|$
  5. Draw a conclusion: $\begin{cases} p\text{-value} \leq \alpha \text{ (where } \alpha \text{ is the level, .05 by default)} \implies \text{reject } H_0 \\ p\text{-value} > \alpha \implies \text{retain } H_0 \text{ as plausible} \end{cases}$
- A $(100\%)(1 - \alpha)$ confidence interval for μ is $\bar{X} \pm t_{n-1,\alpha/2} \frac{S}{\sqrt{n}}$.

The relationship between a two-sided test and a confidence interval

Recall our 95% CI for μ_{paint}, $\bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \approx 1.348 \pm .181 = (1.167, 1.529)$. A CI is a range of ____________________ for μ in light of the data. Our interval contains $\mu_0 = 1.5$, so $H_0 : \mu = 1.5$ is _________, and we would _____________ H_0. More generally, these two statements are equivalent:

- A level-α test of $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$ ____________________ H_0 (because it's _____________ that $\mu = \mu_0$, in light of the sample mean $\bar{x}$).
- μ_0 falls _____________ a $1 - \alpha$ confidence interval for μ (a range of _____________ values for μ, in light of the sample mean $\bar{x}$).

Here's a picture from §5 (well, I added "H_0 :", changed μ to μ_0, σ to S, and $z_{\alpha/2}$ to $t_{n-1,\alpha/2}$):



On the other hand, for an $\bar{x}$ in the tails, we _________ H_0 and μ_0 would be _________ the CI.

e.g. Two-sided CI vs. test

1. A researcher interested in the aluminum recycling market collects 20 cans that he regards as a simple random sample from the local population of cans. He finds a 99% confidence interval for μ, the population mean weight of cans, as $14.2 \pm .04$ g.

   (a) What decision should he make about $H_0 : \mu = 14.3$ vs. $H_A : \mu \neq 14.3$ at level $\alpha = .01$?

   (b) What decision should he make about $H_0 : \mu = 14.22$ vs. $H_A : \mu \neq 14.22$ at level $\alpha = .01$?

2. The P-value for a two-sided test of $H_0 : \mu = 10$ vs. $H_A : \mu \neq 10$ is 0.06.

   (a) Does the 95% confidence interval for μ include 10? Why?

   (b) Does the 90% confidence interval for μ include 10? Why?

Power (for the known-σ case)

Recall:

- $\beta = P(\text{type \_\_\_\_\_\_ error}) = P(\text{do not reject } H_0 | H_0 \text{ is false})$
- power $= 1 - \beta = P(\text{reject } H_0 | H_0 \text{ is false})$

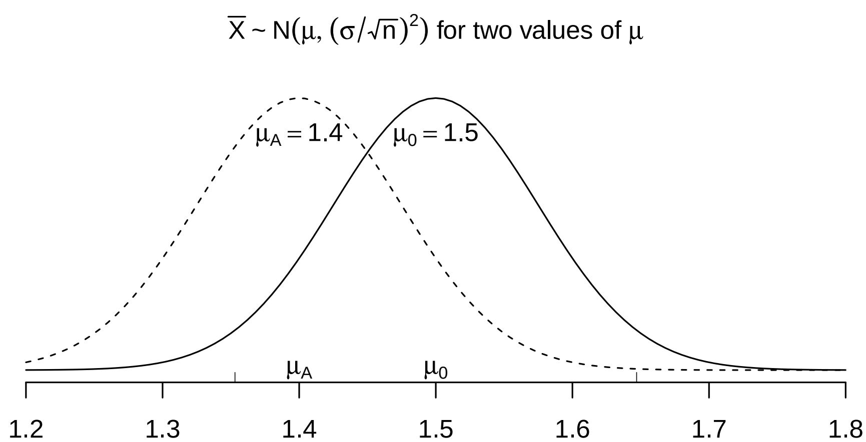
Neither is well-defined until we choose a particular value, ______, in the region specified by H_A.

e.g. For the paint test of $H_0 : \mu = 1.5$ vs. $H_A : \mu \neq 1.5$, suppose we know $\sigma_{\text{paint thickness}} = 0.30$ mil. Find power$_{\mu_A = 1.4}$.

1. Use $H_0 : \mu = 1.5$ and $\alpha = .05$ to find the rejection region: ________________________.

   By unstandardizing from $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$, find the equivalent rejection region in $\bar{X}$ as

   $\bar{X} <$ ____________ or $\bar{X} >$ ____________ (draw)



2. Now use the particular H_A value μ_A to find power$_{\mu_A = 1.4}$ = ____________

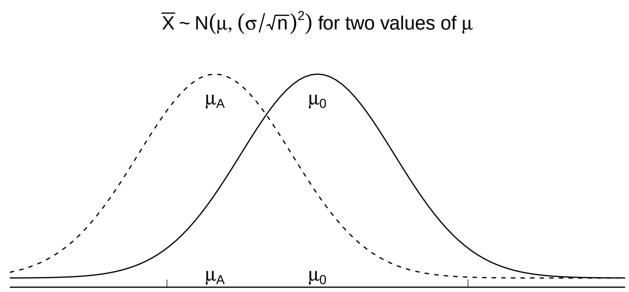
This power is ___________: if the true mean were $\mu = 1.4$, we would probably _______________ H_0 based on a sample of size 16, even though H_0 is _______.

To increase power,

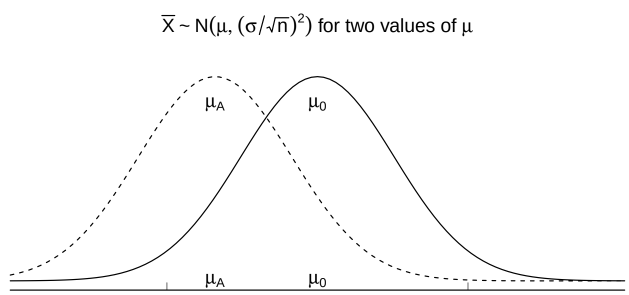
- ______________ $|\mu_0 - \mu_A|$
- ______________ the type-I error rate, α
- ______________ the sample size, n
- ______________ the population standard deviation, σ

Here we find formulas to do similar power calculations more generally.

- For a one-sided test, $H_A : \mu < \mu_0$ (or $H_A : \mu > \mu_0$), power = ______________



- For a two-sided test, $H_A : \mu \neq \mu_0$ power = ______________



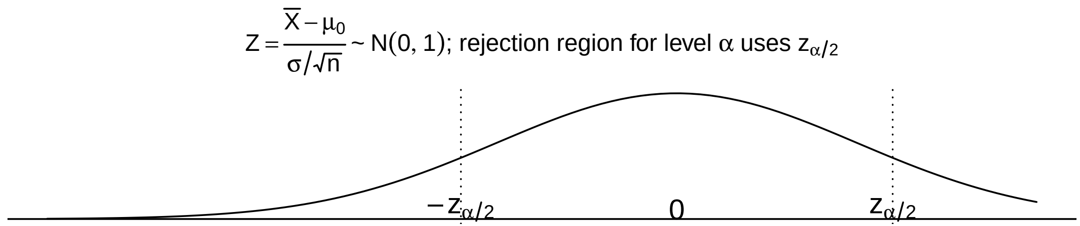
Power of a test of $H_0 : \mu = \mu_0$ when H_0 is false because $\mu = \mu_A$:

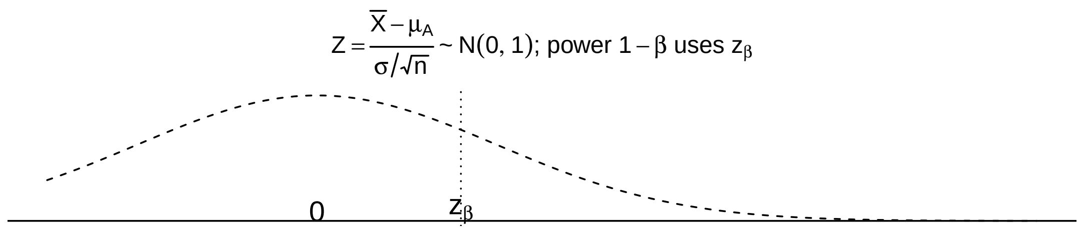
- For a one-sided test, $H_A : \mu < \mu_0$ or $H_A : \mu > \mu_0$, $\text{power}_{\mu_A} = P\left(Z < \frac{|\mu_0 - \mu_A|}{\sigma/\sqrt{n}} - z_\alpha\right)$.
- For a two-sided test, $H_A : \mu \neq \mu_0$, $\text{power}_{\mu_A} \approx P\left(Z < \frac{|\mu_0 - \mu_A|}{\sigma/\sqrt{n}} - z_{\alpha/2}\right)$.

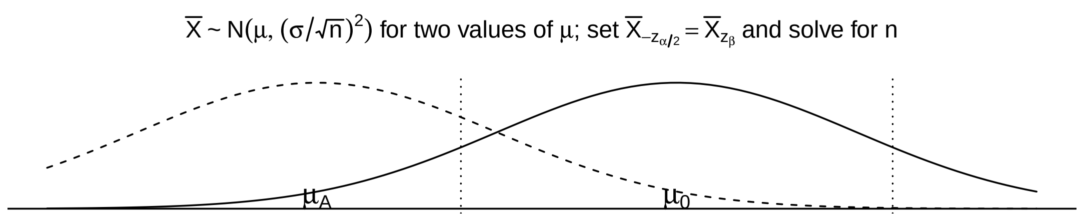
e.g. Check that the formula gives the same power for the paint test as we found earlier.

Power and sample size

Now we find the sample size n required to achieve power $1-\beta$ to reject H_0 at level α when a particular H_A is true:







For a test of $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$ at level α, the sample size n required to have power $1-\beta$ when the true μ is μ_A is $n \approx \left(\frac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu_A}\right)^2$.

e.g. For the paint, suppose $\sigma = 0.30$ is known, and we seek the sample size n required to have power 0.8 to reject H_0 at level $\alpha = .05$ when the true mean is $\mu_A = 1.4$. We need $n =$ ______________

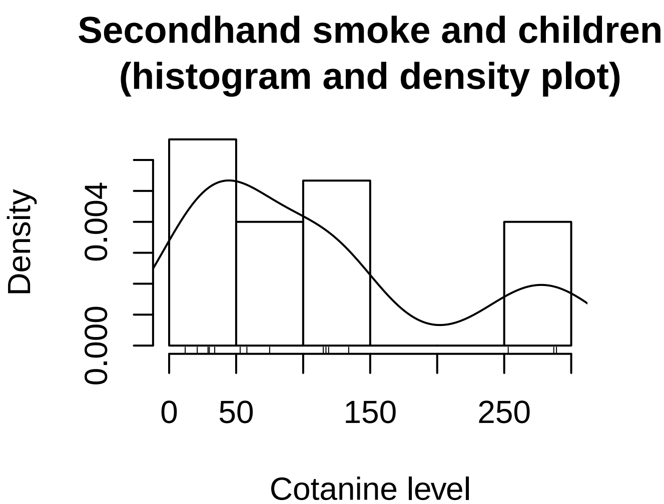
7 More One-Sample Confidence Intervals and Tests, Part 2 of 2

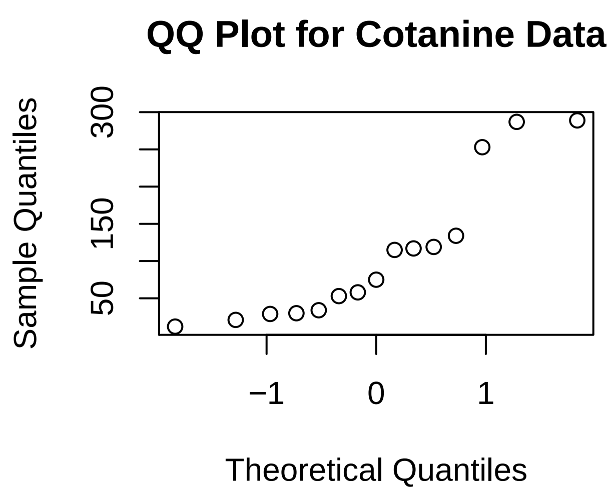
Bootstrap for a confidence interval or test for μ

So far, our discussion of estimating the population mean μ has assumed either the population is normal, so that $\bar{X}$ is also ______________, or the sample size is ______________________ for the CLT to indicate that $\bar{X}$ is approximately normal. What if neither is true?

e.g. Secondhand smoke presents health risks, especially to children. A SRS was taken of 15 children exposed to secondhand smoke, and the amount of cotanine (a metabolite of nicotine) in their urine was measured. The data were: 29, 30, 53, 75, 34, 21, 12, 58, 117, 119, 115, 134, 253, 289, 287. Are these data strong evidence the mean cotanine is above 75 units in kids exposed to secondhand smoke? (It is below 75 in unexposed kids.)

First, check graphs to see whether an assumption of a normal population is ______________ :





This looks pretty bad, so we worry about a normality assumption. The sample is small, so the CLT may not help. Without a normal $\bar{X}$, the quantity

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

will not have a _____________________ distribution. The *bootstrap* is a sneaky way to estimate the true distribution of this T. It estimates the ____________________________ of a statistic by sampling with replacement from a simple random sample from a popluation. e.g. Here's a hand-waving account . . .

To use the bootstrap to make a confidence interval or do a hypothesis test for a mean μ,

1. Collect one simple random sample of size n from the population. Compute the sample mean, $\bar{x}$ (an estimate of the population mean, μ) and the sample standard deviation, s (an estimate of the population standard deviation, σ).

2. Draw a random sample of size n, ________________________, from the data. Call these observations x_1^*, x_2^*, ..., x_n^*. Some data may appear more than once in this resampling, and some not at all.

3. Compute the ____________ and ______ of the resampled data. Call these $\bar{x}^*$ and s^*.

4. Compute the statistic $\hat{t} = \dfrac{\bar{x}^* - \bar{x}}{s^*/\sqrt{n}}$

5. Repeat steps 2-4 a large number of times, accumulating many $\hat{t}$'s. They approximate the (unknown) sampling distribution of $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$.

6. To find a $(100\%)(1-\alpha)$ confidence interval for μ, find the $1-\alpha/2$ and $\alpha/2$ upper *critical values* of the approximate sampling distribution, calling them $\hat{t}_{(1-\alpha/2)}$ and $\hat{t}_{(\alpha/2)}$. The bootstrap $100(1-\alpha)\%$ confidence interval is $\left(\bar{x} - \hat{t}_{(\alpha/2)}\dfrac{s}{\sqrt{n}}, \bar{x} - \hat{t}_{(1-\alpha/2)}\dfrac{s}{\sqrt{n}}\right)$.

7. To test $H_0 : \mu = \mu_0$, compute $t_{\text{obs}} = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$. Find the p-value, an area under the approximate sampling distribution density curve given by ____________, where m depends on H_A:

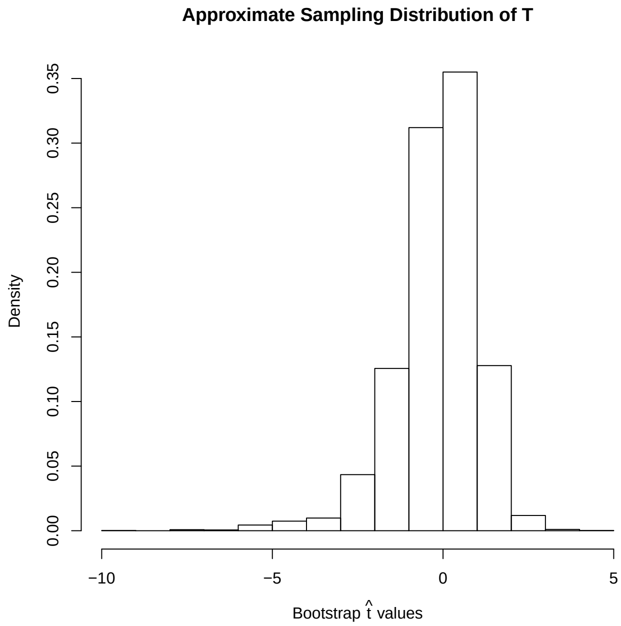
   $H_A : \mu > \mu_0 \implies m$ is the number of values of $\hat{t}$ for which $\hat{t}$______t_{obs}

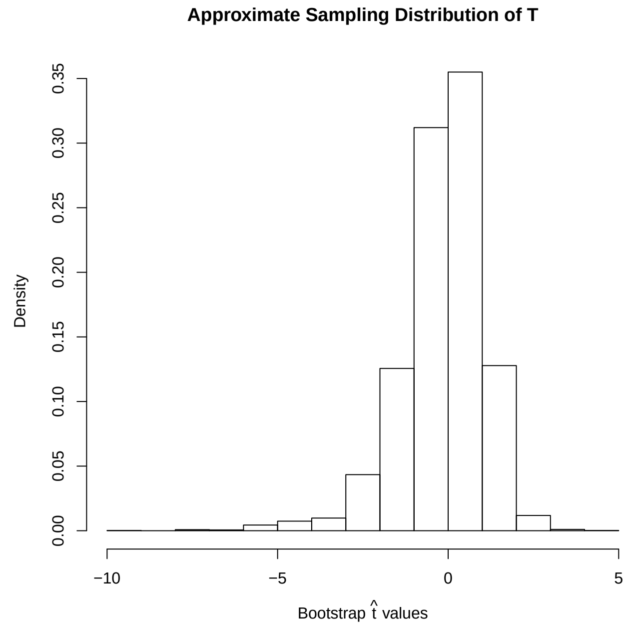
   $H_A : \mu < \mu_0 \implies m$ is the number of values of $\hat{t}$ for which $\hat{t} < t_{\text{obs}}$

   $H_A : \mu$ _ $\mu_0 \implies m$ is the number of values of $\hat{t}$ for which $\hat{t} < -|t_{\text{obs}}|$ or $\hat{t} > |t_{\text{obs}}|$

   Draw a conclusion as usual: $\begin{cases} p\text{-value} \leq \alpha \text{ (where } \alpha \text{ is the level, .05 by default)} \implies \text{reject } H_0 \\ p\text{-value} > \alpha \implies \text{retain } H_0 \text{ as plausible} \end{cases}$

e.g. For the secondhand smoke data, we find $\bar{x} = 108.4$ and $s = 95.6$. Bootstrapping 5000 times yields the following approximate distribution of t (draw for interval on left and for test on right):





Unlike a t or normal distribution, this distribution is ____________ symmetric.

Make a bootstrap confidence interval for μ = population mean cotanine in smoky kids

The upper critical values, from R, are $\hat{t}_{(1-\alpha/2)} = -3.56$ and $\hat{t}_{\alpha/2} = 1.86$ (draw, above left), so the interval is

$\left(108.4 - (\_\_\_\_\_\_\_\_\_)\frac{95.6}{\sqrt{15}}, 108.4 - (\_\_\_\_\_\_\_\_\_)\frac{95.6}{\sqrt{15}}\right) \approx (62.5, 196.3)$.

This interval is not symmetric about $\bar{x}$. It would __________________ on bootstrapping again.

Run a bootstrap test for μ

We wish to know whether μ is greater than 75, so we test $H_0 : \mu = 75$ vs. H_A : ____________.

Find t_{obs} = ____________.

Draw the p-value, above right. ___

Here (from R) $m = 348$ of the bootstrap values were greater than 1.353, so the p-value is __________________, and, at level $\alpha = .05$, we would __________________ H_0.

Here is one way to do this bootstrap using R:

```
# Create a new function, bootstrap(x, n.boot), having two inputs:
#   - x is a data vector
#   - n.boot is the desired number of resamples from x
# It returns a vector of n.boot t-hat values.
bootstrap = function(x, n.boot) {
  n = length(x)
  x.bar = mean(x)
  t.hat = numeric(n.boot) # create vector of length n.boot zeros
  for(i in 1:n.boot) {
    x.star = sample(x, size=n, replace=TRUE)
    x.bar.star = mean(x.star)
    s.star = sd(x.star)
    t.hat[i] = (x.bar.star - x.bar) / (s.star / sqrt(n))
  }
  return(t.hat)
}

# Use the bootstrap() function to get an approximate sampling
# distribution of T for the smoke data.
data = c(29, 30, 53, 75, 34, 21, 12, 58, 117, 119, 115, 134, 253, 289, 287)
B = 5000
t.hats = bootstrap(data, B)

# Plot the approximate sampling distribution.
hist(t.hats, freq=FALSE, xlab = "Bootstrap t-hat values",
     main = "Approximate Sampling Distribution of T")

n = length(data) # Get summary statistics.
x.bar = mean(data)
s = sd(data)
cat(sep="", "n=", n, ", x.bar=", x.bar, ", s=", s, "\n")

# Make a CI for mu. First find quantiles for a 95% interval.
t.lower = quantile(t.hats, probs=.025) # This is our t_{1 - alpha.2}.
t.upper = quantile(t.hats, probs=.975) # This is our t_{alpha/2}.
cat(sep="", "t.lower=", t.lower, ", t.upper=", t.upper, "\n")
ci.low  = x.bar - t.upper * s / sqrt(n) # This is our lower interval endpoint.
ci.high = x.bar - t.lower * s / sqrt(n) # This is our upper interval endpoint.
cat(sep="", "confidence interval: (", ci.low, ", ", ci.high, ")\n")

# Run a test of H_0: mu = m_0. First find t_{obs}.
mu.0 = 75
t.obs = (x.bar - mu.0) / (s / sqrt(n))
```

```
cat(sep="", "t.obs=", t.obs, "\n")
# sum() counts the TRUE values by first converting TRUE / FALSE values to 1 / 0.
m.left = sum(t.hats < t.obs) # This is for H_A: mu < mu_0.
p.value.left = m.left / B
cat(sep="", "m.left=", m.left, ", B=", B, ", p.value.left=", p.value.left, "\n")
m.right = sum(t.hats > t.obs) # This is for H_A: mu > mu_0.
p.value.right = m.right / B
cat(sep="", "m.right=", m.right, ", B=", B, ", p.value.right=", p.value.right, "\n")

# This is for H_A: mu != mu_0. ("!=" means "is not equal to.")
m.left.abs = sum(t.hats < -abs(t.obs))
m.right.abs = sum(t.hats > abs(t.obs))
p.value.two.sided = (m.left.abs + m.right.abs) / B
cat(sep="", "m.left.abs=", m.left.abs, ", m.right.abs=", m.right.abs,
    ", B=", B, ", p.value.two.sided=", p.value.two.sided, "\n")
```

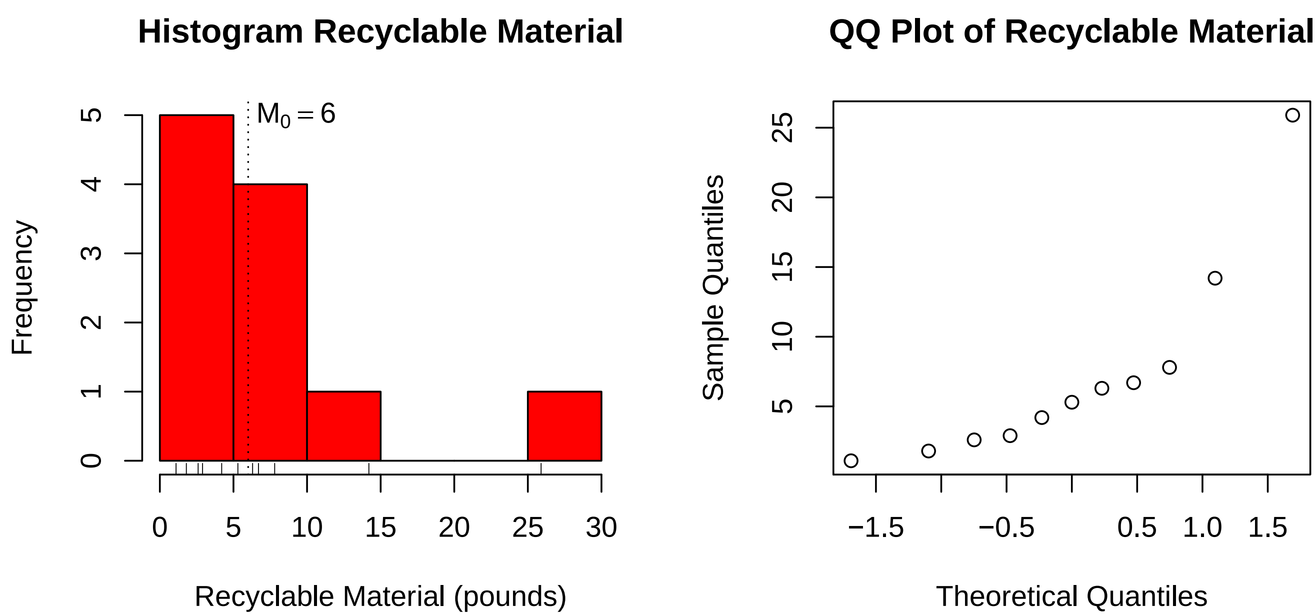
Sign test for an unknown median M

If the data do not seem to be from a normal population and the sample size is small, an alternative to the bootstrap is the sign test. It is a test for a ____________. If the population is roughly ____________, the sign test is equivalent to a test for a ____________.

e.g. A city trash department is considering separating recyclables from trash to save landfill space and sell the recyclables. Based on data from other cities, if more than half the city's households produce 6 lbs or more of recyclable material per collection period, the separation will be profitable. A random sample of 11 households yields these data on material per household in pounds:

$$14.2, 5.3, 2.9, 4.2, 1.8, 6.3, 1.1, 2.6, 6.7, 7.8, 25.9$$

We start with plotting. Here are a histogram and QQ plot:



Neither plot suggests a normal population: both show ________________. Since we have $n = 11$, the CLT is __________, and, in any case, our question is really about a ____________. So, letting M be the population ____________, we test:

$$H_0 : M = 6$$
$$H_A : \text{____________}$$

We need a test statistic. If H_0 is true, the sample should have about ____________ of the observations greater than 6 and ____________ less than 6. The probability of observing a value greater than 6 in the sample should be ______. A natural choice of test statistic is the number, B, of observations greater than 6. Under H_0, $B \sim$ ____________________. (Note: n is the number of observations ____________________ the null value of the median. If any of the observations were equal to 6, we would ____________________.)

The value of the test statistic is $b =$ _____________

(Equivalently, this is the number of positive differences from M_0. These differences are:

$$\_\_\_\_\_\_, \_\_\_\_\_\_, -3.1, -1.8, -4.2, 0.3, -4.9, -3.4, 0.7, 1.8, \_\_\_\_\_\_$$

Of these differences, _______ are positive. The *sign test* counts the number of "+" signs.)

The p-value is _____________

R can find this p-value via `sum(dbinom(x=5:11, size=11, prob=.5))`.

Our conclusion is:

For a two-sided test, find $P(B \geq b)$ and $P(B \leq b)$ and use ___________________________.

Summary:

Suppose $X_1, \ldots, X_n$ is a simple random sample from a population with median M. To test that M has a specified value, M_0,

0. (First discard any data equal to M_0, reducing n accordingly.)
1. State null and alternative hypotheses, $H_0 : M = M_0$ and H_A
2. Check assumptions
3. Find differences from the median, $X_1 - M_0, \ldots, X_n - M_0$, and the test statistic, $B =$ number of positive differences
4. Find the p-value, which is a probability for $B \sim \text{Bin}(n, .5)$ depending on H_A:

   $H_A : M > M_0 \implies p\text{-value} = P(B \geq b)$

   $H_A : M < M_0 \implies p\text{-value} = P(B \leq b)$

   $H_A : M \neq M_0 \implies p\text{-value} = \text{minimum}\{2P(B \leq b), 2P(B \geq b), 1\}$

5. Draw a conclusion: $\begin{cases} p\text{-value} \leq \alpha \text{ (where } \alpha \text{ is the level, .05 by default)} \implies \text{reject } H_0 \\ p\text{-value} > \alpha \implies \text{retain } H_0 \text{ as plausible} \end{cases}$

Estimation of an unknown population proportion π

e.g. An accounting firm has a large list of clients (the population), with an information file on each client. The firm has noticed errors in some files and wishes to know the proportion of files that contain an error. Call the population proportion of files in error π. An SRS of size $n = 50$ is taken and used to estimate π. Now the firm will decide whether it is worth the cost to examine and fix all the files. Each file sampled was classified as containing an error (call this 1), or not (call this 0). The results are:

Files with an error: 10; files without errors, 40.

To develop an estimator of π, recall the binomial distribution: $X \sim Bin(n, \pi)$ is the ____________________ in ______ independent trials, each having ____________ possible outcomes (success and failure), and each having probability ______ of success. We found $E(X) =$ ______, $VAR(X) =$ ________.

Our estimator of the population proportion is the *sample proportion* $P = -$ Here are some of its properties:

- $E(P) =$

- $VAR(P) =$

- $SD(P) = \sqrt{\frac{\pi(1-\pi)}{n}}$

This tells us our estimator P is ____________ for π, and gives a measure of precision. As in the discussion of $\bar{X}$, we can estimate the standard deviation by plugging in our estimator of π:

To make a confidence interval or do a test for π, we need the distribution of P. Its exact distribution is related to the binomial distribution, which is difficult to use in this context. However, the CLT can help. If n is large enough, the conditions of the CLT are met, because $X = \sum Y_i$ (where Y_i is a Bernoulli trial, either 0 or 1), so $P = \frac{X}{n} = \frac{1}{n} \sum Y_i$ is a __________________. . Thus, for large samples, P is approximately ____________ distributed:

$$P \sim N\left(\pi, \left[\sqrt{\frac{\pi(1-\pi)}{n}}\right]^2\right) (\approx)$$

We want to use this distribution to make a confidence interval for π and do a test on π, but we don't know π, so we have to estimate the standard deviation, $\sqrt{\frac{\pi(1-\pi)}{n}}$.

- For the interval, use the sample proportion P to estimate π, so the standard deviation of P is about $S_P = \sqrt{\frac{P(1-P)}{n}}$. A rule of thumb says we need the numbers of successes and failures, _______ and _______, each to be greater than 5 for the CLT approximation to be reasonable.

  The $100\%(1-\alpha)$ confidence interval for π is then $P \pm z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$.

  Proof:

  e.g. Find a 95% CI for the unknown proportion π of defective files.

- The test comes with a null hypothesis, $H_0 : \pi = \pi_0$, so we should use π_0 for π in the standard deviation of P and say, if H_0 is true, then $P \sim N\left(\pi, \left[\sqrt{\frac{\pi_0(1-\pi_0)}{n}}\right]^2\right)$ $(\approx)$. A rule of thumb says we need the expected numbers of successes and failures, _______ and _______, each to be greater than 5 for the CLT approximation to be reasonable. Standardizing gives $Z = \frac{P-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \sim N(0,1)$, which we can use as a test statistic.

  e.g. The CEO decides that if $\pi > .12$, it will be worthwhile to review and fix every file. Run a test to help the CEO decide.

  We test: $H_0 : \pi = (\pi_0 = 0.12)$ vs. $H_A :$ _____________.

  In our example we have $n\pi_0 =$ ____________________ and $n(1-\pi_0) =$____________________.

  We observed $z_{obs} =$ _____________

  Our p-value = ___________________________

  Our conclusion is ___________________________

Summary:

Let X be the number of successes in a large number n of independent Bernoulli trials, each having probability π of success. Let $P = \frac{X}{n}$.

- To test that π has a specified value, π_0, where $n\pi_0 > 5$ and $n(1-\pi_0) > 5$,

  1. State null and alternative hypotheses, $H_0 : \pi = \pi_0$ and H_A
  2. Check assumptions
  3. Find the test statistic $Z = \dfrac{P - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$
  4. Find the p-value, which depends on H_A:
     $H_A : \pi > \pi_0 \implies p\text{-value} =$ $P(Z > z)$, the area right of z
     $H_A : \pi < \pi_0 \implies p\text{-value} =$ $P(Z < z)$, the area left of z
     $H_A : \pi \neq \pi_0 \implies p\text{-value} =$ $P(Z < -|z|) + P(Z > |z|)$, the sum of the ares left of $-|z|$ and right of $|z|$
  5. Draw a conclusion: $\begin{cases} p\text{-value} \leq \alpha \text{ (where } \alpha \text{ is the level, .05 by default)} \implies \text{reject } H_0 \\ p\text{-value} > \alpha \implies \text{retain } H_0 \text{ as plausible} \end{cases}$

- An approximate $100\%(1-\alpha)$ confidence interval for π is $P \pm z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$, provided $X > 5$ and $n - X > 5$.

Demonstrate that $P \sim N(\ldots)$

Our CI and test for π relied on the CLT to say $P = \frac{X}{n} \sim N(\ldots)(\approx)$ because P is a sample mean. X is a sample sum, which is also $N(\ldots)$. In particular, $X \sim \text{Bin}(n, \pi) \approx N(n\pi, n\pi(1-\pi))$.

Here is a graphical comparison of $\text{Bin}(n, \pi)$ with $N(n\pi, n\pi(1-pi))$ for $n = 20$ and several values of π to help with understanding the CLT claim and our rule-of-thumb requiring $n\pi > 5$ and $n(1-\pi) > 5$. (You may ignore the code. I'll run it and discuss it.)

```
n=20
delta.p=.1
for (p in seq(from=delta.p, to=1-delta.p, by=delta.p)) {
  Sys.sleep(3)
  y=dbinom(x=0:n, size=n, prob=p)
  curve(dnorm(x, mean=n*p, sd=sqrt(n*p*(1-p))), 0, n, ylab="",
        main=bquote("n=" * .(n) * ", " * pi * "=" * .(p) *
          ", " * n * pi * "=" * .(n*p) * ", " * n * (1- pi) *
          "=" * .(n*(1-p))))
  segments(x0=0:n, y0=0, y1=y)
}
```

In §8, we compare _____ populations via independent samples.

Extra examples (if time allows)

Extra confidence intervals for μ with known or unknown σ

The basal diameter of a sea anemone indicates its age. Suppose the population mean (μ) and standard deviation (σ) are unknown.

1. Here are the diameters of a simple random sample of 40 anemones: 4.3, 5.7, 3.9, 4.8, 3.5, 3.5, 1.3, 4.6, 4.4, 3.7, 4.9, 5.6, 5.1, 2.3, 2.3, 6.9, 5.4, 3.6, 4.3, 4.1, 3.2, 4.6, 2.8, 4.9, 4.5, 4.4, 5.8, 3.6, 5.6, 2.6, 1.5, 4.1, 4.7, 6.5, 5.4, 3.8, 3.4, 4.9, 5.5, 7.2. These data have $\bar{x} = 4.33$ and $s = 1.329$. Find a 95% confidence interval for μ or explain why you cannot.

2. Here is a simple random sample of 12 anemone diameters: 5.3, 2.8, 5.2, 2.9, 2.5, 2.9, 3.0, 2.9, 5.2, 4.3, 3.7, 2.7. Find a 95% confidence interval for μ or explain why you cannot.

3. Here is a simple random sample of 12 anemone diameters: 3.5, 6.5, 3.6, 2.8, 4.2, 4.2, 1.8, 5.7, 2.6, 4.7, 4.9, 4.4. Find a 95% confidence interval for μ or explain why you cannot.

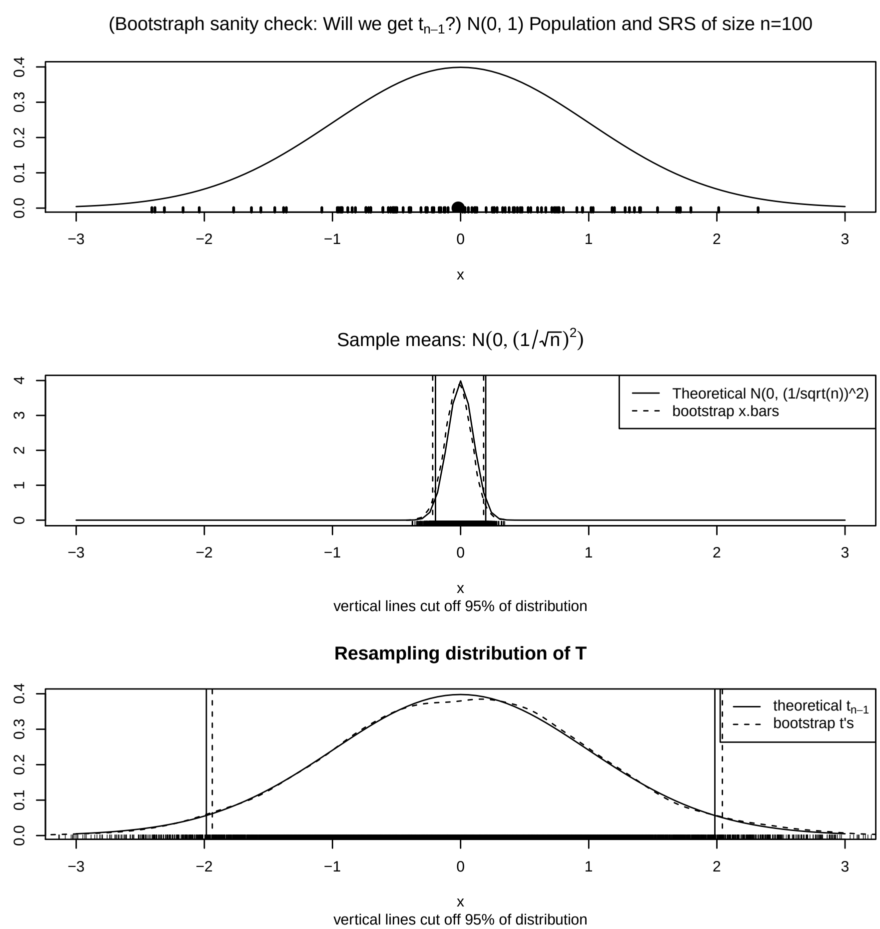
4. (Inference) Now suppose the population mean (μ) is unknown but $\sigma = 1.4$ cm is known. What changes in the intervals above?

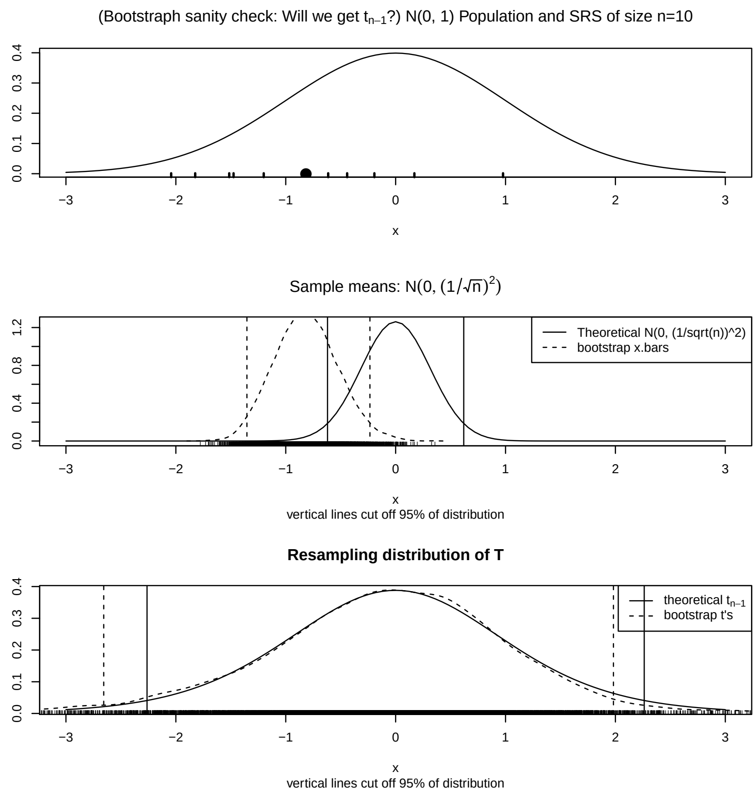
Extra bootstrap sanity check

- Bootstrap confidence interval for an unknown mean

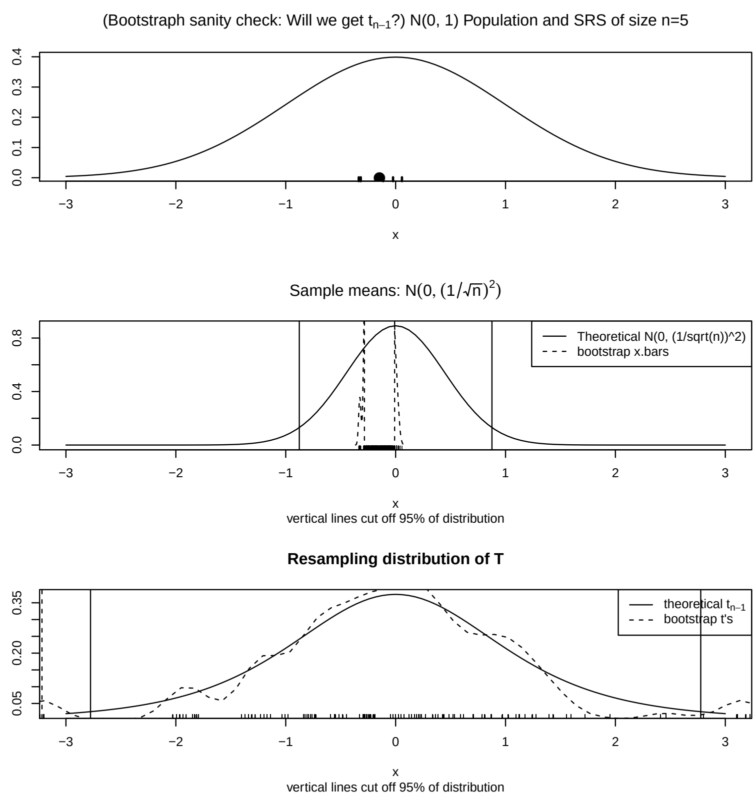
  Let's do "sanity check" computer simulations to see whether the bootstrap does something reasonable when we know what to expect. Suppose $X_1, \ldots, X_n$ is a SRS from $N(0, 1^2)$, a normal population. In this case, we know that $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$.

  - Do a bootstrap to get an approximate sampling distribution (labeled "bootstrap t's" in the graphs) for T and see whether it looks like t_{n-1}.
  - How do the results depend on n?
    * Resampling with replacement from a large sample seems like a good approximation to repeated sampling from the population.
    * Resampling with replacement from a small sample seems like a lousy approximation to repeated sampling from the population.





(Bootstraph sanity check: Will we get t_{n-1}?) N(0, 1) Population and SRS of size n=10
x
Sample means: $N(0, (1/\sqrt{n})^2)$
Theoretical N(0, (1/sqrt(n))^2)
bootstrap x.bars
x
vertical lines cut off 95% of distribution
Resampling distribution of T
theoretical t_{n-1}
bootstrap t's
x
vertical lines cut off 95% of distribution



(Bootstraph sanity check: Will we get t_{n-1}?) N(0, 1) Population and SRS of size n=5
x
Sample means: $N(0, (1/\sqrt{n})^2)$
Theoretical N(0, (1/sqrt(n))^2)
bootstrap x.bars
x
vertical lines cut off 95% of distribution
Resampling distribution of T
theoretical t_{n-1}
bootstrap t's
x
vertical lines cut off 95% of distribution

Extra sign test for M

e.g. A clinical trial measured survival time in weeks for 10 lymphoma patients as 49, 58, 75, 110, 112, 132, 151, 276, 281, 362+, where "+" indicates a patient still alive at the end of the study. Are these data strong evidence the population median survival time M for lymphoma patients is is different than 200?

Extra inference for one proportion π

e.g. Monica learned in first grade that about 71% of Earth's surface is covered in water. To see whether this made sense, she asked her brother to toss her a spinning inflatable globe 100 times. For 66 of her catches, her right pointer finger tip was on water, while for 34 it was on land. Now she's stuck. Help her by finding and interpreting a 99% confidence interval for the proportion of Earth covered by water in light of her data.

e.g. Do children prefer vanilla or chocolate ice cream? To test this, a teacher gave a random sample of 33 students the choice. 24 of 33 chose chocolate, and the other 9 chose vanilla. Use these data to test the hypothesis that, in the population, students have no preference.

Here is a supplement to §7 on hypothesis testing and power. It's p. 10.5 or p. 25 in the §7 notes.

Suppose the engine painter sells engines to General Motors (GM), promising mean thickness $\mu = 1.5$ for a truckload of engines. GM doesn't want to use the engines if $\mu = 1.4$ because they will rust too quickly. (GM wants the engines for $\mu = 1.5$ and for very close values; it doesn't want them for values far from 1.5. This page focuses on the fact that GM doesn't want them if $\mu = 1.4$.)

An independent lab measures the paint thickness for a random sample of n of the engines. It tests $H_0 : \mu = 1.5$ vs. $H_A : \mu \neq 1.5$. There are four possible outcomes:

	reject H_0	do not reject H_0
H_0 is true because $\mu = \mu_0$	P(type I Error) $= \alpha$	P(correct) $= 1 - \alpha$
H_0 is false because $\mu = \mu_A$	P(correct) $=$ power$_{\mu_A} = (1 - \beta_{\mu_A})$	P(type II Error) $= \beta_{\mu_A}$

- Suppose the engine painter ships a good truckload with $\mu = 1.5$.
  - $\alpha = P(\text{type I error}) = P(\text{reject } H_0 | H_0 \text{ is true})$ is the engine painter's risk, due to an unlucky sample, of having the lab say it's a bad truckload, so the engine painter does not get paid. The engine painter wants α _____________.
  - $1 - \alpha = P(\text{do not reject } H_0 | H_0 \text{ is true})$ is probability of the lab result matching reality. GM pays for and uses correctly-painted engines.
- Suppose the engine painter ships a bad truckload with $\mu = 1.4$.
  - $\beta_{\mu=1.4} = P(\text{type II error}) = P(\text{do not reject } H_0 | H_0 \text{ is false because } \mu = 1.4)$ is GM's risk, due to an unlucky sample, of having the lab say there's no strong evidence the truckload is bad, so GM pays for and uses the bad truckload. GM wants $\beta_{\mu=1.4}$ _____________.
  - power$_{\mu=1.4} = 1 - \beta_{\mu=1.4} = P(\text{reject } H_0 | H_0 \text{ is false because } \mu = 1.4)$ is the probability of the lab result matching reality. GM does not pay for or use bad engines.
- There is tension between the painter's desire for low α and GM's desire for low β. A contract could specify a lower α and a higher β in exchange for GM paying the painter _______; or a higher α and a lower β would require GM to pay the painter _______.
- The conflict between α and β can be resolved by increasing the sample size.

  $H_A : \bar{X} \sim N(\mu_A, \sigma^2/n)$ $\qquad$ $H_0 : \bar{X} \sim N(\mu_0, \sigma^2/n)$