One-Sample Tests

One-sample t-test

Recall example from “5 Estimation” notes:

- e.g. A car manufacturer ... warehouse ... thousands of painted blocks ... 16 blocks are selected at random, and the paint thickness is measured ...:
  1.29, 1.12, 0.88, 1.65, 1.48, 1.59, 1.04, 0.83, 1.76, 1.31, 0.88, 1.71, 1.83, 1.09, 1.62, 1.49

Suppose the specification says the thickness should be 1.50 mil. We want to know whether the device is off this mark on average, so that the machine should be re-calibrated to correct its population mean thickness, \( \mu \). Our first step is to graph the data:

In light of running `qqnorm(rnorm(16))` several times, these plots leave it plausible that the population is __________. We use the hypotheses, \( H_0 : \text{__________ mil vs. } H_A : \text{__________ mil.} \)

If the population is normal or \( n \) is large enough that the CLT applies, and if \( H_0 \) is true, then \( \bar{X} \sim N(\mu_0, \sigma^2/n)(\approx) \). Then \( T = \frac{\bar{X} - \mu_0}{S \sqrt{n}} \sim t_{n-1} \). (\( \mu_0 \) is the value of \( \mu \) under \__________: here \( \mu_0 = \__________ \).

Values of \( \bar{X} \) far from \( \mu_0 \) (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 $\alpha = P(\text{type I error}) = \ldots$. Compute $t$ and compare it to the region.

  This region is $T < -t_{n-1, \alpha/2}$ or $T > t_{n-1, \alpha/2}$, the complement of the interval $-t_{n-1, \alpha/2} < T < t_{n-1, \alpha/2}$ (draw).

For the paint, we have $n = 16$, so we need $t_{16-1, 0.025} = \ldots$. Our rejection region is \ldots. Our observed $t$ is $t_{obs} = \bar{x} - \mu_0 = \ldots$

Conclusion:

- Compute the p-value and compare it to $\alpha$.

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

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

  - from the $t$ table, \ldots, or
  - from R, use \(2 * \text{pt}(q = -1.796, \ df=15)\) = \ldots

Conclusion:

- Compute a confidence interval and check whether $\mu_0$ is in it (below).
Suppose $X_1, \cdots, X_n$ is a simple random sample from a normal population with mean $\mu$ and standard deviation $\sigma$, or $n$ is large. To test that $\mu$ 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, $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

4. Find the $p$-value, which is an area under $t_{n-1}$ depending on $H_A$:
   - $H_A : \mu > \mu_0 \implies p$-value $= P(T > t)$, the area right of $t$ (where $T \sim t_{n-1}$)
   - $H_A : \mu < \mu_0 \implies p$-value $= P(T < t)$, the area left of $t$
   - $H_A : \mu \neq \mu_0 \implies p$-value $= P(|T| > |t|)$, the sum of areas left of $-|t|$ and right of $|t|$

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

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

Recall the 95% CI for $\mu$ we calculated in §5: $\bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \approx 1.348 \pm .180 = (1.168, 1.528)$. A CI is a range of values for $\mu$ in light of the data. Our interval contains $\mu_0 = 1.5$ mil, so $H_0 : \mu = 1.5$ mil is $\underline{\text{unplausible}}$, and we would $\underline{\text{reject}}$ $H_0 : \mu = 1.5$ mil.

More precisely, these two statements are equivalent:

- A level-\(\alpha\) test of $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$ $\underline{\text{reject}}$ $H_0$ (because it’s $\underline{\text{unplausible}}$ that $\mu = \mu_0$, in light of the sample mean $\bar{x}$).
- $\mu_0$ falls $\underline{\text{inside}}$ a $1 - \alpha$ confidence interval for $\mu$ (a range of $\underline{\text{plausible}}$ values for $\mu$, in light of the sample mean $\bar{x}$).

Here’s a picture repeated from §5 (well, I added “$H_0$ :” and changed $\mu$ to $\mu_0$):

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

On the other hand, for a value of $\bar{x}$ out in the shaded tails, we would $\underline{\text{reject}}$ $H_0$ and $\mu_0$ would be $\underline{\text{outside}}$ the confidence interval around $\bar{x}$. 
e.g. In §5 we found an unknown-σ $t$-interval for $\mu$ from the paint data as $1.35 \pm 0.18$. (We also found a known-σ $z$-interval as $1.35 \pm 0.15$.) What does it say about the test of $H_0 : \mu = 1.5$?

e.g. 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 $\mu$ include 10? Why?

b. Does the 90% confidence interval for $\mu$ include 10? Why?
Power (for the known-σ case)

Recall:

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

Neither is well-defined until we choose a particular value, $\square$, 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 $\text{power}_{\mu_A = 1.4}$.

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

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

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

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

- \( |\mu_0 - \mu_A| \)
- the type-I error rate, \( \alpha \)
- the sample size, \( n \)
- the population standard deviation, \( \sigma \)

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 = \( \frac{\mu - \mu_0}{\sigma/\sqrt{n}} + z_\alpha \), for two values of \( \mu \)

- For a two-sided test, \( H_A : \mu \neq \mu_0 \) power = \( \frac{\mu - \mu_0}{\sigma/\sqrt{n}} + z_{\alpha/2} \), for two values of \( \mu \)

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 \( H_A : \mu > \mu_0 \), power_{\mu_A} = \( P\left(Z < \frac{|\mu_0 - \mu_A|}{\sigma/\sqrt{n}} - z_\alpha \right) \).
- For a two-sided test, \( H_A : \mu \neq \mu_0 \), power_{\mu_A} \approx \( P\left(Z < \frac{|\mu_0 - \mu_A|}{\sigma/\sqrt{n}} - z_{\alpha/2} \right) \).
Power and sample size

Now we find the sample size required to achieve power $1 - \beta$ to reject $H_0$ at level $\alpha$ when a particular $H_A$ is true. This can only be done iteratively when $\sigma$ is unknown. But if we assume $\sigma$ is known, then we can use $N(0, 1)$ rather than $t_{n-1}$ and can find $n$:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1);$$

rejection region for level $\alpha$ uses $z_{\alpha/2}$

$$Z = \frac{\bar{X} - \mu_A}{\sigma / \sqrt{n}} \sim N(0, 1);$$

power $1 - \beta$ uses $z_\beta$

$$\bar{X} \sim N(\mu, (\sigma / \sqrt{n})^2)$$

for two values of $\mu$; set $\bar{X} - z_{\alpha/2} = \bar{X}_b$ and solve for $n$

For a test of $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$ at level $\alpha$, the sample size $n$ required to have power $1 - \beta$ when the true $\mu$ is $\mu_A$ 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 $H_0$ when the true mean is $\mu_A = 1.4$. We need $n =$ ____________
Test using bootstrap

What if the data cannot be assumed to come from a normal distribution and \( n \) is too small for the CLT? We will see two tests for this situation. The first is based on the bootstrap. We return to the §5 secondhand smoke example. Recall:

e.g. Secondhand smoke ... risks ... to children. A SRS ... 15 children exposed ... amount of cotanine in ... urine ... measured. Cotanine ... should be below 75 units.

The data were:

29, 30, 53, 75, 34, 21, 12, 58, 117, 119, 115, 134, 253, 289, 287

We would start by checking a QQ plot, which we have already done. It didn’t support an assumption of a normal population, and since \( n = 15 \) is not large, we probably shouldn’t use the CLT. As in the CI setting, we proceed with a bootstrap. Since we wish to know whether the average cotanine level is greater than 75 units, we let \( \mu \) be the population mean cotanine level, and then test \( H_0 : \mu = 75 \) vs. \( H_A : \mu > \mu_0 \).

This is a \( \text{one-sided} \) test, since the alternative goes in only one direction.

To do a bootstrap test for an \( H_0 : \mu = \mu_0 \),

1. Collect one simple random sample of size \( n \) from the population. Compute the sample mean, \( \bar{x} \), which is an estimate of the population mean, \( \mu \), the sample standard deviation, \( s \), and \( t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \).

2. Draw a random sample of size \( n \), \( x_1^*, x_2^*, ..., x_n^* \), 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 \( \bar{x}^* \) and \( s^* \) of the resampled data. Call these \( \bar{x}^* \) and \( s^* \).

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

5. Repeat steps 2-4 a large number, \( B \), times, accumulating many \( \hat{t} \)'s. They approximate the sampling distribution of \( T \).

6. Find the \( p \)-value, which is an area under the approximate sampling distribution density curve given by \( \text{by} \), where \( m \) depends on \( H_A \):

\[
H_A : \mu > \mu_0 \implies m \text{ is the number of values of } \hat{t} \text{ for which } \hat{t} \geq t_{\text{obs}}
\]

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

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

7. Draw a conclusion:

\[
\{ p\text{-value} \leq \alpha \text{ (where } \alpha \text{ is the level, .05 by default) } \implies \text{ reject } H_0 \}
\]

\[
p\text{-value} > \alpha \implies \text{ retain } H_0 \text{ as plausible}
\]
Doing this for the secondhand smoke data, we found $\bar{x} = 108.4$ and $s = 95.6$, so $t_{obs} = \phantom{0000}$.

Bootstrapping $B = 5000$ times yields the following approximate distribution of $t$:

![Approximate Sampling Distribution of t](image)

Here (from R) $m = 348$ of the bootstrap values were greater than 1.353, so the p-value is \phantom{0000}, and, at level $\alpha = .05$, we would \phantom{0000} $H_0$. However, the p-value is relatively small and suggests there is some evidence that population mean cotanine level is greater than 75 units.

For the one-sided alternate hypothesis in the other direction,

$$H_0 : \mu = 75$$
$$H_A : \mu < 75,$$

our p-value and conclusion are

For the two-sided alternate hypothesis,

$$H_0 : \mu = 75$$
$$H_A : \mu \neq 75,$$

we would need to find the number of bootstrap samples where $\hat{t} < -1.353$, and the number where $\hat{t} > 1.353$. We then take the sum of those two numbers and divide by $B$. In our case, these numbers (from R) are 348 and 700, so the p-value and conclusion are
Here is one way to do this bootstrap using R:

```r
# (The bootstrap() function, below, is the same as the one in # the estimation.pdf notes.)

# 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)
}

smoke = c(29, 30, 53, 75, 34, 21, 12, 58, 117, 119, 115, 134, 253, 289, 287)

# Find t_{obs}.  
# Find t_{obs}.
 n = length(smoke)
 x.bar = mean(smoke)
 s = sd(smoke)
 mu.0 = 75
 t.obs = (x.bar - mu.0) / (s / sqrt(n))

# Use the bootstrap() function to get an approximate sampling # distribution of T for the smoke data.
B = 5000
smoke.boot <- bootstrap(smoke, B)

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

summary(smoke.boot > t.obs)
# sum() counts the TRUE values by first converting TRUE / FALSE values to 1 / 0.
m = sum(smoke.boot > t.obs) # This ">") depends on H_A.
p.value = m / B
print(p.value)
```
Sign test

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:

$$H_0 : M = 6$$

$$H_A : \quad$$

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 $H_0$, $B \sim \text{Bin}(n, 0.5)$. (Note: $n$ is the number of observations equal to 6, we would _______.)

The value of the test statistic is $b =$ _______.

(Equivalently, this is the number of positive differences of data from the median assumed under $H_0$. These differences are:

_____, _____, -2.1, -0.8, -3.8, -1.7, -3.9, -2.4, 1.7, 2.8, _____

Of these differences, ______ are positive. The sign test counts the number of “+” signs.)

The p-value is ___________

R can do this sign test via `sum(dbinom(x=5:11, size=11, prob=.5))` or `binom.test(x=5, n=11, p=.5, alternative="greater")`

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, $M_0$,

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

5. Draw a conclusion: 
   
   \[
   \begin{align*}
   \text{p-value} \leq \alpha \text{ (where } \alpha \text{ is the level, .05 by default)} \implies \text{ reject } H_0 \\
   \text{p-value} > \alpha \implies \text{ retain } H_0 \text{ as plausible}
   \end{align*}
   \]
Test for one population proportion

Recall the accounting example from §5: A firm has an information file on each of a large number of clients. Call the population proportion of files with errors \( \pi \). The CEO decides that if \( \pi > .12 \), it will be worthwhile to review and fix every file. An SRS of size \( n = 50 \) is taken with the same result as before:

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

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

Recall our point estimate for \( \pi \), the sample proportion \( P = \frac{X}{n} = \) \_

Use the CLT, as in §5, to say that under \( H_0 \),

\[
P \sim N \left( \pi_0, \frac{\pi_0(1 - \pi_0)}{n} \right) = N \left( 0.12, \frac{0.12(1 - 0.12)}{50} \right) \approx N(0, 1)
\]

which implies

\[
Z = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} = \frac{P - 0.12}{\sqrt{\frac{0.12(1 - 0.12)}{50}}} \sim N(0, 1)
\]

(We need the expected numbers of successes and failures each to be greater than 5 for the CLT approximation to be reasonable. In our example we have \( n\pi_0 = \) \_

\( n(1 - \pi_0) = \) \_

We observed \( z_{obs} = \) \_

Our p-value is 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 $\pi$ of success. Let $P = \frac{X}{n}$. To test that $\pi$ has a specified value, $\pi_0$, 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 $Z = \frac{P - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$

4. Find the $p$-value, which depends on $H_A$:
   - $H_A: \pi > \pi_0 \implies p$-value = $P(Z > z)$, the area right of $z$
   - $H_A: \pi < \pi_0 \implies p$-value = $P(Z < z)$, the area left of $z$
   - $H_A: \pi \neq \pi_0 \implies p$-value = $P(|Z| > |z|)$, the sum of the areas left of $-|z|$ and right of $|z|$

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

Extra examples

Sign test

e.g. Here is a SRS of 20 component lifetimes (in hours):

1.7, 3.3, 5.1, 6.9, 12.6, 14.4, 16.4, 24.6, 26.0, 26.5, 32.1, 37.4, 40.1, 40.5, 41.5, 72.4, 80.1, 86.4, 87.5, 100.2

Are these data strong evidence that the population median lifetime exceeds 25.0 hours?
Test for one proportion

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.

In §8, we compare _____ independent populations.