One Sample Hypothesis Tests

Ex: Suppose a bag of 32 cookies has a mean cookie weight of 10.9 g. Suppose we know
std dev. to be 0.5 g. Is it reasonable that
this is random variation from a population mean of 11g?

Hypothesis Test ⇒ test a statistical hypothesis

6 Steps:
1) State hypotheses (always in terms of population
parameters j Steps 1-3 in book)
2) Determine the level of significance
(.05 unless otherwise specified)
3) Determine the test statistic
(something that can be looked up in a table)
and its sampling distribution (which table)
4) Compute the test statistic and either the
critical region or the p-value.
5) Reject or fail to reject the null hypothesis
6) State conclusions in the context of the original problem
Ex: $\bar{x} = 10.9$, $n = 32$, $\sigma = 0.5$

1) Claim $\mu = 11$, population mean
   versus $\mu < 11$

   The one with the equality is the null hypothesis, the other one
   is the alternative hypothesis.

   \[ H_0: \mu = 11 \]  where $\mu$ is the population mean cookie weight
   \[ H_1: \mu < 11 \]

   The null is the default. We only conclude the alternative if
   there is enough evidence.

   Like in a trial where the defendant is innocent until proven guilty.

   Failure to reject is a lack of evidence. It does not mean the
   null is necessarily true.

   If we are trying to prove something, it must be the alternative.
   We can see if the null is reasonable.

   If we show the null is not reasonable, $H_0: \mu \geq 11$
   we conclude the alternative is highly likely to be true.

2) Level of significance is $\alpha = 0.05$ unless specified otherwise

3) Test statistic

   Here we are testing a mean, with a known $\sigma$

   Test statistic is $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

   If $\sigma$ were unknown, use $t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ t dist/n

   For a proportion, $Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$ normal

   $\frac{p - P}{\sqrt{P(1-P)/n}}$

4) The critical region is the set of values of
   the test statistic that would cause you to reject the null
   hypothesis. Values that would be highly unusual if the
   null were true. "Unusual" is defined by the level of
   significance - the probability we are willing to be wrong
   when the null is true.

   If the null is true, then $\frac{X - \mu}{\sigma/\sqrt{n}} = Z$ follows the standard normal
   distribution. If this is too small, we will reject the null.

   \[
   \frac{\bar{X}}{\sigma/\sqrt{n}}
   \]

   From Z table $P(Z < -1.645) = 0.05$

   So we reject the null hypothesis if $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < -1.645$
So our critical region is $z < -1.645$

The actual test statistic is $z = \frac{10.9 - 11}{.5/\sqrt{18}} = -1.13$

5) Fail to reject $H_0$ because $-1.13 \neq -1.645$

6) Fail to reject the claim that the population mean cookie weight is 11g.
   
   or: Conclude it is reasonable that the population mean cookie weight is 11g.

**Type I and Type II errors**

A Type I error is rejecting the null when it is true
   (convicting an innocent person)

A Type II error is failing to reject when the null is false
   (not convicting a guilty person)

In general, we consider a Type I error to be worse,
   so we limit those to a fixed significance level ($\alpha$),
   and then try to minimize the probability of a Type II error.

The power of a test is the probability of rejecting when the alternative is true.
   $Power = 1 - P(\text{Type II error})$

**Ex 2:** A manufacturer has a soda-filling machine and they are concerned that it may not be properly calibrated.

A sample of 18 20oz. bottles is found to have an average content of 19.96oz. with a std. dev. of 0.04oz. Is it reasonable that the machine is properly calibrated?

1) Claim: $\mu = 20$ or $\mu \neq 20$

   $H_0: \mu = 20$  
   $H_1: \mu \neq 20$

   where $\mu$ is the population mean bottle contents

2) $\alpha = .05$

3) Test for a mean with $\sigma$ unknown, test statistic is $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

   Sampling distribution is $t$ with 17 df
   (degrees of freedom $= n - 1 = 18 - 1 = 17$)

4) Two-tailed reject if $t < -2.11$ or $t > 2.11$

   $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{19.96 - 20}{.04/\sqrt{18}} = -4.24$
P-values

An alternative to finding the critical region is to compute the p-value, a measure of how unusual the observed data would be if the null hypothesis is true.

\[ p\text{-value} = \text{probability of observing a test statistic as or more extreme than that actually observed when the null hypothesis is true} \]

(\text{it is a conditional probability})

We reject the null when \( p\text{-value} < \alpha \) (typically \( \alpha < .05 \))

**Ex 1:** \( H_0: \mu = 11 \) vs. \( H_i: \mu \leq 11 \) (one sided)

reject if \( z < -1.645 \) so "extreme" is very negative

\[ z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = -1.13 \]

\( p\text{-value} = P(Z < -1.13) = .1292 \)

.1292 > .05 so fail to reject \( H_0 \)

If \( \mu = 11 \), the probability of seeing a test statistic less than or equal to -1.13 is .1292.

**Ex 2:** \( H_0: \mu = 20 \) vs. \( H_i: \mu \neq 20 \)

reject if \( t < -2.11 \) or \( t > 2.11 \) "extreme" is either very negative or very positive

\[ t = -4.24 \]

\[ p = P(t < -4.24) + P(t > 4.24) = 2 \times P(t < -4.24) \]

From table, less than .01 in twotails \( p < .01 \)

so reject \( H_0 \)

From JMP \( p = .00055 \)