Hypothesis tests for a mean

Starter

1. (a) If $X_1, X_2, X_3, \ldots, X_n$ is a random sample from N$(\mu, 1)$, state the distribution of the sample mean $\bar{X}$.
   (b) Find the least sample size required to ensure that the probability that $\bar{X}$ is within 0.1 of μ is greater than 0.95.

Notes

A hypothesis test can be carried out on a sample to see whether the sample could come from the population. This can be used to check whether an intervention strategy has worked.

For example, imagine the mean length of time to recover from a cold is 10 days with a standard deviation of 2.5 days. A company develops a drug that it hopes will reduce the time to recovery. It tests the drug on 100 people who have colds and measures their time to recovery. The mean time to recover of this sample is 8.9 days.

On the face of it, it appears that the new drug works but is the value of 8.9 days statistically significant? A statistician for the drug company will ask the question "could this sample come from a population whose mean is 10 days and whose standard deviation is 2.5 days?". The company will hope the answer is "no" because this would show there has been a reduction in time.

Structure of a hypothesis test

The hypothesis test investigates whether the mean of the sample, the ***test statistic*** $\bar{x}$, is consistent with coming from a population with the given mean.

The significance level, $\alpha\ \%$, indicates what level of doubt the statistician is willing to allow. A test statistic that "passes" a hypothesis test at the $1\ \%$ significance level provides greater evidence than one that passes at the $5\ \%$ level.

1. State the null hypothesis: it is always $H_0 : \mu = \mu_0$ where μ_0 is the given population mean
2. Decide whether you will carry out a $1-$tailed or $2-$tailed test (it is usually given in the question).
3. State the alternative hypothesis, H_1:
   - $H_1 : \mu \neq \mu_0$ — two-tailed test
   - $H_1 : \mu > \mu_0$ — one-tailed test, it is believed the sample mean comes from a population with a higher mean.
   - $H_1 : \mu < \mu_0$ — one-tailed test, it is believed the sample mean comes from a population with a lower mean.
4. Decide on the level of significance (it is often given in the question, usually $5\ \%$)
5. Either find the ***$p-$value corresponding to the test statistic*** or determine the ***critical value***, $\bar{x}_{cv}$, ***corresponding to*** the level of significance.
6. Either:
   (a) compare the $p-$value to the level of significance
   - If p − value > significance level $\Rightarrow$ do not reject H_0
   - If p − value < significance level $\Rightarrow$ reject H_0
   - ***N.B.*** If the test is two-tailed, it is p − value $> \dfrac{\text{significance level}}{2}$

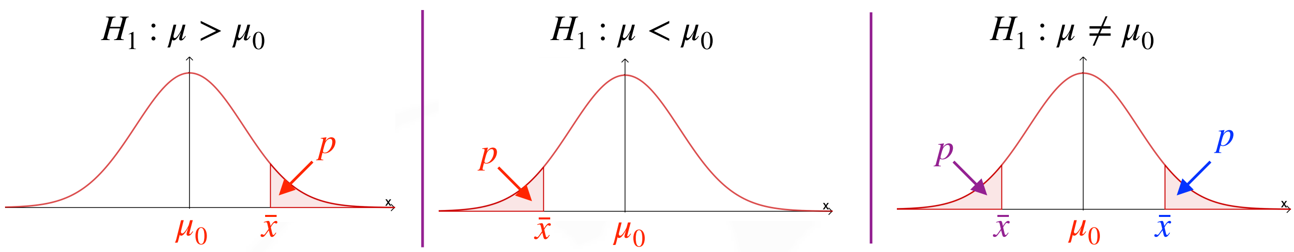
   (b) compare the $x-$value for the significance to the test statistic. This $x-$value is the ***critical value*** and defines the ***critical region***.
   - If the $x-$value does not lie in the critical region $\Rightarrow$ reject H_0
   - If the $x-$value lies in the critical region $\Rightarrow$ reject H_0

Using $p-$ values in hypothesis tests

Here, the test statistic, $\bar{x}$, defines the $p-$value.

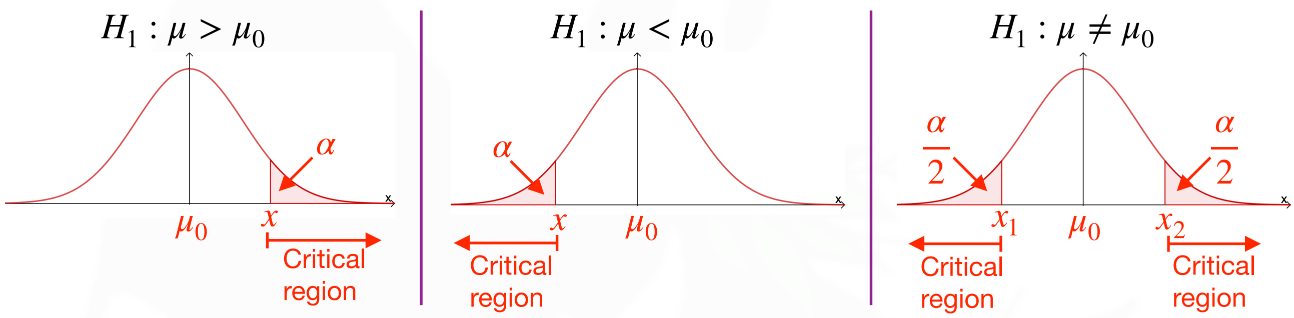
If the $p-$value is less than the level of significance, α % , we reject H_0.

With two-tailed tests, calculate the $p-$value depending on which side of μ_0 the test statistic, $\bar{x}$, lies.



Critical regions and critical values in hypothesis tests

If the test statistic, $\bar{x}$, lies in the critical region, we reject H_0.



To determine the critical region, calculate the critical value, x, such that:

$P(X > x) = \alpha$ $\qquad$ $P(X > x) = \alpha$ $\qquad$ $P(X < x_1) = \frac{\alpha}{2}$ or $P(X > x_2) = \frac{\alpha}{2}$

E.g. The mean length of time to recover from a cold is 10 days with a standard deviation of 2.5 days. A company develops a drug that it says reduces the time to recovery. It tests the drug on 36 people who have colds and measures their time to recovery. The mean time to recover of this sample is 9.4 days. Test at the 5 % level whether the drug is effective.

Working: $X \sim \text{N}(10, 2.5^2) \quad \Rightarrow \quad \bar{X}_{36} \sim \text{N}\left(10, \frac{2.5^2}{36}\right)$

Null hypothesis, $H_0 : \mu = 10$

Alternative hypothesis, $H_1 : \mu < 10$

$n = 100, \bar{x} = 9.4$, level of significance is 5 %

$p-$ value method

$P(\bar{X} < 9.4) = 0.0749 \equiv 7.49$ %

Since 7.49 % > 5 % , $\bar{x} = 9.4$ ***does not lie*** in the critical region. Therefore, we ***do not reject*** H_0 and conclude that there is evidence to suggest the drug is not effective in reducing recovery time.

Critical value method

Let x_{cv} be the critical value such that $P(\bar{X} < x_{cv}) = 0.05 \Rightarrow x_{cv} = 9.31$

The critical value is 9.31 and the recovery time for the sample must be less than this value for it to be significant.

Since $\bar{x} = 9.4 \not< 9.31 = x_{cv}$, we do not reject H_0 and conclude that there is evidence to suggest the drug is not effective in reducing recovery time.

Notice that the phrase “there is evidence to suggest” because it may be that the drug is effective but not on this sample. In fact, had the sample size been bigger, say 100, with the same sample mean of 9.4, the company’s statistician could have rejected H_0 i.e. the drug does reduce recovery time. Alternatively, by using a significance level of $10\,\%$, the sample mean value of 9.4 would have meant the rejection of H_0.

E.g. 1 Use the “$p-$value” method to test these hypotheses at the stated levels:

(a) $H_0 : \mu = 5, H_1 : \mu \neq 5, n = 25, \bar{x} = 6.1, \sigma = 3.0$ at the $5\,\%$ level
(a) $H_0 : \mu = 100, H_1 : \mu < 100, n = 36, \bar{x} = 98.5, \sigma = 5$ at the $5\,\%$ level
(a) $H_0 : \mu = 15, H_1 : \mu > 15, n = 40, \bar{x} = 16.5, \sigma = 3.5$ at the $1\,\%$ level

Working: (a) $X \sim \text{N}(5, 3.0^2) \quad \Rightarrow \quad \bar{X}_{25} \sim \text{N}\left(5, \dfrac{3.0^2}{25}\right)$

Since $\bar{x} = 6.1$ is greater than $\mu = 5$, calculate $P(\bar{X} > 6.1)$.
$p = P(\bar{X} > 6.1) = 0.0334 \equiv 3.34\,\%$
Since it is a two-tailed test, $\alpha = \dfrac{5\,\%}{2} = 2.5\,\%$
Since $p = 3.34\,\% > 2.5\,\%$, we ***do not reject*** H_0 at the $5\,\%$ level.

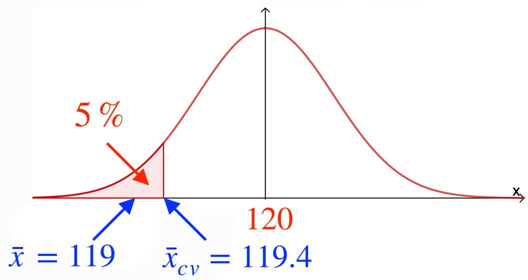
E.g. 2 Use the “critical region” method to test these hypotheses at the stated levels:

(a) $H_0 : \mu = 120, H_1 : \mu < 120, n = 30, \bar{x} = 119, \sigma = 2.0$ at the $5\,\%$ level.
(b) $H_0 : \mu = 12.5, H_1 : \mu > 12.5, n = 25, \bar{x} = 12.9, \sigma = 1.5$ at the $1\,\%$ level.
(c) $H_0 : \mu = 0, H_1 : \mu \neq 0, n = 45, \bar{x} = 0.9, \sigma = 3.0$ at the $5\,\%$ level.

Draw a separate diagram for each question.

Working: (a) $X \sim \text{N}(120, 2.0^2) \quad \Rightarrow \quad \bar{X}_{30} \sim \text{N}\left(120, \dfrac{2.0^2}{30}\right)$

Let $\bar{x}_{cv}$ be the critical value such that $P(\bar{X} < \bar{x}_{cv}) = 0.05$
$\Rightarrow \bar{x}_{cv} = 119.4$
Since $\bar{x} = 119 < 119.4 = \bar{x}_{cv}$, the ***test statistic lies in the critical region*** so we ***reject*** H_0.



N.B. When you do worded questions make sure that you write H_0 and H_1 clearly.

E.g. 3 The IQ scores of a population are normal distributed with a mean of 100 and standard deviation of 15. A psychologist wishes to test the theory that eating chocolate before sitting an IQ test improves your score. A random sample of 81 people are selected and they are each given a 50 g bar of chocolate to eat before taking a standard IQ test. Their mean score on the test was 103.1. Test the psychologist’s theory at the $2\,\%$ level.

E.g. 4 The mean weight of men in a particular country is found to have a standard deviation of 9.1 kg. The sample mean based on a random sample of 120 men was found to be 75.3 kg. A medical journal claims that the mean weight of men in this country is 77.1 kg. Test the validity of this statement at the $5\,\%$ level.

Video: An introduction to hypothesis testing
Video: Mean of Normal distribution hypothesis testing
Video: Sample means: hypothesis testing example

Solutions to Starter and E.g.s

Exercise
p405 18B Qu 1i, 2i, 3i, 4-8, (9 red)

Summary
Success criteria — hypothesis test for a sample mean:

1. State the null hypothesis: it is always $H_0 : \mu = \mu_0$ where μ_0 is the given population mean
2. Decide whether you will carry out a $1-$tailed or $2-$tailed test (it is usually given in the question).
3. State the alternative hypothesis, H_1:
   - $H_1 : \mu \neq \mu_0$ — two-tailed test
   - $H_1 : \mu > \mu_0$ — one-tailed test, it is believed the sample mean comes from a population with a higher mean.
   - $H_1 : \mu < \mu_0$ — one-tailed test, it is believed the sample mean comes from a population with a lower mean.
4. Decide on the level of significance (it is often given in the question, usually $5\,\%$)
5. Either find the ***$p-$value corresponding to the test statistic*** or determine the ***critical value***, $\bar{x}_{cv}$, ***corresponding to*** the level of significance.
6. Either:
   (a) compare the $p-$value to the level of significance
   - If p − value $>$ significance level $\Rightarrow$ do not reject H_0
   - If p − value $<$ significance level $\Rightarrow$ reject H_0
   - ***N.B.*** If the test is two-tailed, it is p − value $> \dfrac{\text{significance level}}{2}$

   (b) compare the $x-$value for the significance to the test statistic. This $x-$value is the ***critical value*** and defines the ***critical region***.
   - If the $x-$value does not lie in the critical region $\Rightarrow$ reject H_0
   - If the $x-$value lies in the critical region $\Rightarrow$ reject H_0