Some other standard hypothesis tests

***Z*-test for a normal mean**

Context: $Y_1,...,Y_n \sim iid\ N(\mu,\sigma^2)$, where σ^2 is known

H_0: $\mu = \mu_0$
H_a: $\mu \neq \mu_0$
TS: $Z = \dfrac{\bar{Y} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$ (if H_0 is true)
RR: $|Z| > z_{\alpha/2}$ (Note that this test is exact, unlike the last two.)

Example 7

A bottling machine dispenses volumes that are normally distributed with a variance of 16 square ml. 8 bottles were sampled and their average volume was 941.6 ml.

Test whether the machine dispenses 950 ml on average, at the 5% level.

H_0: $\mu = 950$
H_a: $\mu \neq 950$
TS: $Z = \dfrac{\bar{Y} - 950}{4/\sqrt{8}} \sim N(0,1)$
RR: $|Z| > z_{0.025} = 1.96$

$$z = \frac{941.6 - 950}{4/\sqrt{8}} = -5.94$$

|–5.94| > 1.96. Thus $z \in RR$. So reject H_0.

We conclude that the machine does *not* dispense 950 ml on average.

In many situations the population variance is *unknown*. In that case the above hypothesis test cannot be used, and the appropriate test is the following one.

t-test for a normal mean

Context: $Y_1,...,Y_n \sim iid\ N(\mu,\sigma^2)$, where σ^2 is unknown

$H_0:$ $\mu = \mu_0$

$H_a:$ $\mu \neq \mu_0$

TS: $T = \dfrac{\bar{Y} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$ under H_0

RR: $|T| > t_{\alpha/2}(n-1)$

(Note that this test is exact. Also, if *n* is large we can use $z_{\alpha/2}$ instead of $t_{\alpha/2}(n-1)$.)

Example 8

A bottling machine dispenses volumes that are normally distributed with unknown variance.

8 bottles were sampled and their average volume was 941.6 ml.
Also, the sample variance of the 8 volumes was 15.2.

Test whether the machine dispenses 950 ml on average, at the 5% level.

$H_0:$ $\mu = 950$

$H_a:$ $\mu \neq 950$

TS: $T = \dfrac{\bar{Y} - 950}{S/\sqrt{8}} \sim t(7)$

RR: $|T| > t_{0.025}(7)$ = 2.365

$t = \dfrac{941.6 - 950}{\sqrt{15.2}/\sqrt{8}}$ = –6.09

|–6.09| > 2.365. Thus $z \in RR$. So reject H_0.

We conclude that the machine does *not* dispense 950 ml on average (same as in Example 6).

In many situations, it is not reasonable to assume that the sample observations are normally distributed. In that case we should use the following hypothesis test.

***Z*-test for the mean of a distribution**

Context: $Y_1,...,Y_n \sim iid\ (\mu, \sigma^2)$, where *n* is large and σ^2 is unknown

H_0: $\mu = \mu_0$
H_a: $\mu \neq \mu_0$
TS: $Z = \dfrac{\bar{Y} - \mu_0}{S/\sqrt{n}} \stackrel{\cdot}{\sim} N(0,1)$ (by the central limit theorem)
RR: $|Z| > z_{\alpha/2}$

Note: If σ^2 happens to be known, substitute σ for S in *Z*.

All the above tests involve a single population mean μ. However we may want to compare *two* means. In that case it may be appropriate to conduct the following test.

t-test for the difference between two normal means

Context: $X_1,...,X_n \sim iid\ N(\mu_1, \sigma_1^2)$ (1st sample)
$Y_1,...,Y_m \sim iid\ N(\mu_2, \sigma_2^2)$ (2nd sample)
The two samples are independent
σ_1^2 and σ_2^2 are unknown but equal

H_0: $\mu_1 - \mu_2 = \delta$

H_a: $\mu_1 - \mu_2 \neq \delta$

TS: $T = \dfrac{(\bar{X} - \bar{Y}) - \delta}{S_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \sim t(n+m-2)$

where $S_p^2 = \dfrac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$ (the pooled variance)

RR: $|T| > t_{\alpha/2}(n+m-2)$

Example 9

Six bottles filled by a bottling machine today have volumes

1.80, 1.90, 1.75, 1.81, 1.74, 1.82.

Four bottles filled by the same machine yesterday have volumes

1.61, 1.79, 1.78, 1.83.

Using a significance level of 2%, test that the mean volumes were the same today and yesterday. Assume that volumes are normal with the same variance on both days.

We will apply the above t-test, with:

$\delta = 0$, $\alpha = 0.02$

$n = 6$ (1st sample size; today)

$\bar{x} = (1/6)(1.80 + ... + 1.82) = 1.8033$ (1st sample mean)

$s_1^2 = (1/5)\{(1.80^2 + ... + 1.82^2) - 6(1.8033^2)\} = 0.0033067$

(1st sample variance)

$m = 4$ (2nd sample size; yesterday)

$\bar{y} = (1/4)(1.61 + ... + 1.83) = 1.7525$ (2nd sample mean)

$s_2^2 = (1/3)\{(1.61^2 + ... + 1.83^2) - 4(1.7525^2)\} = 0.0094917$

(2nd sample variance)

$s_p^2 = \dfrac{5(0.0033067) + 3(0.0094917)}{8} = 0.0056260$ (pooled variance)

$t_{\alpha/2}(n + m - 2) = t_{0.01}(8) = 2.896$.

So the test is as follows:

H_0: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 \neq 0$

TS: $T = \dfrac{\bar{X} - \bar{Y}}{S_p\sqrt{\dfrac{1}{6} + \dfrac{1}{4}}} \sim t(8)$

RR: $|T| > 2.896$

$$t = \frac{(1.8033 - 1.7525) - 0}{\sqrt{0.0056260}\sqrt{\dfrac{1}{6} + \dfrac{1}{4}}} = 1.05$$

t lies between –2.896 and 2.896. So we do not reject the null hypothesis.

We conclude that there is no difference between the mean volume dispensed by the machine today and the mean volume dispensed by it yesterday.

In some cases the two population variances are possibly unequal and/or known.
In that case we consider the following.

Two variants of the t-test for the difference between two normal means

1. Suppose that σ_1^2 and σ_2^2 are known. Then take:

TS: $$Z = \frac{(\bar{X} - \bar{Y}) - \delta}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0,1)$$

RR: $|Z| > z_{\alpha/2}$

2. Suppose that: **(i)** σ_1^2 and σ_2^2 are unknown and possibly unequal
(ii) the underlying distributions are possibly non-normal
(iii) *n* and *m* are both large.

Then take:

TS: $$Z = \frac{(\bar{X} - \bar{Y}) - \delta}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \mathrel{\dot\sim} N(0,1)$$

RR: $|Z| > z_{\alpha/2}$

In both cases, the "t-test" should instead be called a "Z-test".

Example 10

450 people were sampled in Sydney and their incomes were determined.
The mean and standard deviation of the sample incomes were 12054 and 1501 ($'s).
875 people were sampled in Melbourne and their incomes were determined.
The mean and standard deviation of the sample incomes were 9043 and 989.
Test at the 2% level whether the mean income in Sydney
is exactly $2000 more than the mean income in Melbourne.

H_0: $\mu_1 - \mu_2 = 2000$, H_a: $\mu_1 - \mu_2 \neq 2000$

TS: $$Z = \frac{\bar{X} - \bar{Y} - 2000}{\sqrt{\frac{S_1^2}{450} + \frac{S_2^2}{875}}} \mathrel{\dot\sim} N(0,1),$$ *RR*: $|Z| > z_{0.01} = 2.326$

$$z = \frac{(12054 - 9043) - 2000}{\sqrt{\frac{1501^2}{450} + \frac{989^2}{875}}} = 12.9.$$

So reject H_0. The difference is not $2000.

Definition

The ***p*-value**, or **attained significance level**, is the smallest level of significance α for which the observed data indicate that the null hypothesis H_0 should be rejected.

The *p*-value may also be interpreted as:

"the probability of getting a result as or more inconsistent with H_0 as the result actually observed, given that H_0 is true".

Example 11

A random sample of 300 widgets produced in a factory weighs a total of 565.8 kg, and the sample standard deviation of the weights of the 300 widgets is 7.6 kg. We're interested in whether the average weight of all widgets is 1 kg. Carry out an appropriate hypothesis test at the 5% level and report the associated *p*-value.

$H_0: \mu = 1,$ $\qquad H_a: \mu \neq 1$

TS: $Z = \dfrac{\bar{Y} - 1}{S/\sqrt{300}} \overset{\cdot}{\sim} N(0,1)$, $\qquad RR$: $|Z| > 1.96$

$z = \dfrac{(565.8/300) - 1}{7.6/\sqrt{300}} = 2.02 \in RR$, and so we reject H_0

There is sufficient evidence at the 5% level to conclude that the average weight of all widgets is not exactly 1 kg.

p-value $= P(|Z - 0| > |z - 0|) = 2P(Z > 2.02) = 0.0435$.

Observe that $0.0435 \leq 0.05$.

If the *p*-value had been *greater* 0.05, we would *not* have rejected the null hypothesis.

In general, we reject the null hypothesis H_0 if and only if the *p*-value is less than or equal to the significance level of the test, α.

This means that an alternative way to perform a hypothesis test is to compute the *p*-value and compare it to the significance level, α. In that case there is no need to determine the rejection region.

Note that a *p*-value provides more information than an 'ordinary' hypothesis test, in that a 'very small' *p*-value indicates 'very strong' evidence against the null hypothesis.

One-sided tests

All the tests so far mentioned are **two-sided tests**, in the sense that the rejection region consists of two separate intervals (e.g. $RR = (-\infty, -1.96) \cup (1.96, \infty)$).

It is sometimes meaningful to instead conduct a **one-sided test**.

Such a test may be either a **lower-tail test** or an **upper-tail test**.

Example 12

It has been alleged that the proportion of defective bolts produced by a certain factory is greater than 10%. A sample of 1000 bolts was taken, and 118 of these were found to be defective. Conduct an appropriate hypothesis test at the 5% level and calculate the associated *p*-value.

H_0: $p = 0.1$

H_a: $p > 0.1$ (this, the allegation, is an **upper-tail alternative hypothesis**)

TS: $Z = \dfrac{\hat{p} - 0.1}{\sqrt{0.1(0.9)/1000}} \mathrel{\dot\sim} N(0,1)$

RR: $Z > z_{0.05} = 1.645$ (so the **upper-tail rejection region** is $(1.645, \infty)$)

$z = \dfrac{0.118 - 0.1}{\sqrt{0.1(0.9)/1000}} = 1.90 \in RR \Rightarrow$ Reject H_0.

We conclude that the allegation is true.

The *p*-value is $P(Z > 1.90) \approx 0.0287$.

The test just examined is an *upper*-tail test. The following is a *lower*-tail test.

Example 14

There is concern that a chocolate factory is producing chocolate bars which are on average lighter than the advertised 100 grams. Eight bars were sampled from the production line and weighed. The sample weights had a mean of 96.7 and a standard deviation of 2.9 (grams). Conduct an appropriate hypothesis test at the 1% level, and report the *p*-value. State any assumptions made.

H_0: $\mu = 100$

H_a: $\mu < 100$ (this is a **lower-tail alternative hypothesis**)

TS: $T = \dfrac{\bar{Y} - 100}{S/\sqrt{8}} \sim t(7)$

RR: $T < -t_{0.01}(7) = -2.998$

(so the **lower-tail rejection region** is $(-\infty, -2.998)$)

$t = \dfrac{96.7 - 100}{2.9/\sqrt{8}} = -3.22$, which is in the RR, so reject H_0

We conclude that the factory *is* making underweight chocolate bars.

For this test we have assumed that the weights of chocolate bars produced at the factory are normally distributed.

In this case the *p*-value is $p = P(T < -3.22) = P(T > 3.22)$, by symmetry.
Now, statistical tables tell us that $P(T > 2.998) = 0.01$ and $P(T > 3.499) = 0.005$.
So p is between 0.005 and 0.01. We may conclude that $p = 0.0075 \pm 0.0025$.
(That's the best we can do using the given tables.
Using a computer we find that $p = 0.0073$ exactly.)

The relationship between hypothesis testing and confidence estimation

It is often possible to conduct a hypothesis test by examining a confidence interval.

For example, suppose that in the context $Y \sim N(\mu, \sigma^2)$ we test $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$. Using the test statistic $z = (\bar{y} - \mu_0)/(\sigma/\sqrt{n})$ with significance level α, the appropriate 2-sided rejection region is $\{z : |z| > z_{\alpha/2}\}$. So the acceptance region is $\{z : |z| \leq z_{\alpha/2}\}$, meaning that we accept (i.e. do not reject) H_0 iff (if and only if)

$$|z| \leq z_{\alpha/2} \Leftrightarrow \left|\frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}\right| \leq z_{\alpha/2} \Leftrightarrow -z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{y} - \mu_0 \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow -\bar{y} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq -\mu_0 \leq -\bar{y} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \Leftrightarrow \bar{y} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{y} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Thus, we reject H_0 iff μ_0 lies outside the two-sided $1-\alpha$ CI for μ, $\left[\bar{y} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right]$.

Summary: There are three different ways to conduct a hypothesis test:

- **(i)** the ***'standard' or rejection region approach*** whereby we reject the null hypothesis iff the test statistic lies in a suitable rejection region;
- **(ii)** the ***attained significance level approach*** whereby we reject the null hypothesis iff the *p*-value is less than or equal to the significance level, α; and
- **(iii)** the ***confidence estimation approach***, whereby we reject $H_0 : \theta = \theta_0$ iff the hypothesized value, θ_0, lies outside a suitable $1-\alpha$ confidence interval for θ.