5.5 Hypothesis Testing

Given a sample $x_1, ..., x_n$ from a probability model $f(x;\theta)$ depending on parameter θ, we can produce an estimate $\hat{\theta}$ of θ, and in some circumstances understand how $\hat{\theta}$ varies for repeated samples. Now we might want to test, say, whether or not there is evidence from the sample that true (but unobserved) value of θ is not equal to a specified value. To do this, we use estimate of θ, and the corresponding estimator and its sampling distribution, to quantify this evidence.

In particular, we concentrate on data samples that we can presume to have a normal distribution, and utilize the Theorem from the previous section. We will look at two situations, namely **one sample** and **two sample** experiments.

ONE SAMPLE :	Random variables sample observations	$X_1, ..., X_n \sim N(\mu, \sigma^2)$ $x_1, ...x_n$	Possible Models	$\mu = c_1$ $\sigma = c_2$
TWO SAMPLE :	Random variables sample one observations	$X_1, ..., X_n \sim N(\mu_X, \sigma_X^2)$ $x_1, ...x_n$	Possible Models	$\mu_X = \mu_Y$ $\sigma_X = \sigma_Y$
	Random variables sample two observations	$Y_1, ..., Y_n \sim N(\mu_Y, \sigma_Y^2)$ $y_1, ...y_n$		

5.5.1 Hypothesis Testing for Normal data I - The Z-test

Recall that, if $X_1, ..., X_n \sim N(\mu, \sigma^2)$ are the i.i.d. outcome random variables of n experimental trials, then $\bar{X} \sim N\left(\mu, \sigma^2/n\right)$ and $nS^2/\sigma^2 \sim \chi^2_{n-1}$, with $\bar{X}$ and S^2 statistically independent.
Suppose we want to test the **hypothesis** that $\mu = c$, for some specified constant c, (where, for example, $c = 20.0$) is a plausible model; more specifically, we want to test

$$\begin{array}{lll} H_0 & : \mu = c & \text{the } \textbf{NULL} \text{ hypothesis} \\ H_1 & : \mu \neq c & \text{the } \textbf{ALTERNATIVE} \text{ hypothesis} \end{array}$$

[i.e. we want to test whether H_0 is true, or whether H_1 is true]. Now, we know that, in the case of a Normal sample, the distribution of the estimator $\bar{X}$ is Normal, and

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \implies Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

where Z is a **random variable**. Now, when we have observed the data sample, we can calculate $\bar{x}$, and therefore we have a way of testing whether $\mu = c$ is a plausible model; we calculate $\bar{x}$ from $x_1, ..., x_n$, and then calculate

$$z = \frac{\bar{x} - c}{\sigma/\sqrt{n}}.$$

If H_0 is true, and $\mu = c$, then the **observed** z should be an observation from an $N(0,1)$ distribution (as $Z \sim N(0,1)$), that is, it should be near zero with high probability. In fact, z should lie between -1.96 and 1.96 with probability $1 - \alpha = 0.95$, say, as

$$\text{P}[-1.96 \leq Z < 1.96] = \Phi(1.96) - \Phi(-1.96) = 0.975 - 0.025 = 0.95.$$

If we observe z to be outside of this range, then there is evidence that H_0 is **not true**.
Alternatively, we could calculate the probability p of observing a z value that is **more extreme** than the z we did observe; this probability is given by

$$p = \begin{cases} 2\Phi(z) & z < 0 \\ 2(1 - \Phi(z)) & z \geq 0 \end{cases}$$

If p is very small, say $p \leq \alpha = 0.05$, then again. there is evidence that H_0 is **not true**. In summary, we need to assess whether z is a **surprising** observation from an $N(0,1)$ distribution - if it is, then we can **reject** H_0.

5.5.2 Hypothesis testing terminology

There are five crucial components to a hypothesis test, namely

TEST STATISTIC

NULL DISTRIBUTION

SIGNIFICANCE LEVEL, denoted α

P-VALUE, denoted p.

CRITICAL VALUE(S)

In the Normal example given above, we have that

z is the **test statistic**
The distribution of random variable Z if H_0 is true is the **null distribution**
$\alpha = 0.05$ is the **significance level** of the test (we could use $\alpha = 0.01$ if we require a "stronger" test).
p is the **p-value** of the test statistic under the null distribution
The solution C_R of $\Phi(C_R) = 1 - \alpha/2$ ($C_R = 1.96$ above) gives the **critical values** of the test $\pm C_R$.

EXAMPLE : A sample of size 10 has sample mean $\bar{x} = 19.7$. Suppose we want to test the hypothesis

$$\begin{aligned} H_0 &: \mu = 20.0 \\ H_1 &: \mu \neq 20.0 \end{aligned}$$

under the assumption that the data follow a Normal distribution with $\sigma = 1.0$.

We have

$$z = \frac{19.7 - 20.0}{1/\sqrt{10}} = -0.95$$

which lies between the critical values ± 1.96, and therefore we have no reason to reject H_0. Also, the p-value is given by $p = 2\Phi(-0.95) = 0.342$, which is greater than $\alpha = 0.05$, which confirms that we have no reason to reject H_0.

5.5.3 Hypothesis Testing for Normal data II - The T-test

In practice, we will often want to test hypotheses about μ when σ is unknown. We cannot perform the Z-test, as this requires knowledge of σ to calculate the z statistic.

We proceed as follows; recall that we know the sampling distributions of $\bar{X}$ and s^2, and that the two estimators are statistically independent. Now, from the properties of the Normal distribution, if we have independent random variables $Z \sim N(0,1)$ and $Y \sim \chi^2_\nu$, then we know that random variable T defined by

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

has a Student-t distribution with ν degrees of freedom. Using this result, and recalling the sampling distributions of $\bar{X}$ and s^2, we see that

$$T = \frac{\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)s^2/\sigma^2}{(n-1)}}} = \frac{(\bar{X} - \mu)}{s/\sqrt{n}} \sim t_{n-1}$$

and T has a Student-t distribution with $n-1$ degrees of freedom, denoted $St(n-1)$. Thus we can repeat the procedure used in the σ known case, but use the sampling distribution of T rather than that of Z to assess whether the test statistic is "surprising" or not. Specifically, we calculate

$$t = \frac{(\bar{x} - \mu)}{s/\sqrt{n}}$$

and find the critical values for a $\alpha = 0.05$ significance test by finding the ordinates corresponding to the 0.025 and 0.975 percentiles of a Student-t distribution, $St(n-1)$ (rather than a $N(0,1)$) distribution.

EXAMPLE : A sample of size 10 has sample mean $\bar{x} = 19.7.$ and $s^2 = 0.78^2$. Suppose we want to carry out a test of the hypotheses

$$\begin{aligned} H_0 &: \mu = 20.0 \\ H_1 &: \mu \neq 20.0 \end{aligned}$$

under the assumption that the data follow a Normal distribution with σ unknown.

We have test statistic t given by

$$t = \frac{19.7 - 20.0}{0.78/\sqrt{10}} = -1.22.$$

The upper critical value C_R is obtained by solving

$$F_{t_{n-1}}(C_R) = 0.975$$

where $F_{St(n-1)}$ is the c.d.f. of a Student-t distribution with $n-1$ degrees of freedom; here $n = 10$, so we can use the statistical tables to find $C_R = 2.262$, and not that, as Student-t distributions are symmetric the lower critical value is $-C_R$. Thus t lies between the critical values, and therefore we have no reason to reject H_0. The p-value is given by

$$p = \begin{cases} 2F_{t_{n-1}}(t) & t < 0 \\ 2(1 - F_{t_{n-1}}(t)) & t \geq 0 \end{cases}$$

so here, $p = 2F_{t_{n-1}}(-1.22)$ which we can find to give $p = 0.253$; this confirms that we have no reason to reject H_0.

5.5.4 Hypothesis Testing for Normal data II - testing σ.

The Z-test and T-test are both tests for the parameter μ. Suppose that we wish to test a hypothesis about σ, for example

$$\begin{aligned} H_0 &: \sigma^2 = c \\ H_1 &: \sigma^2 \neq c \end{aligned}$$

We construct a test based on the estimate of variance, s_2. In particular, we saw from the Theorem on p.32 that the random variable Q, defined by

$$Q = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

if the data have an $N(\mu, \sigma^2)$ distribution. Hence if we define test statistic q by

$$q = \frac{(n-1)s^2}{c}$$

then we can compare q with the the critical values derived from a χ^2_{n-1} distribution; we look for the 0.025 and 0.975 quantiles - note that the Chi-squared distribution is not symmetric, so we need two distinct critical values.

In the above example, to test

$$\begin{aligned} H_0 &: \sigma^2 = 1.0 \\ H_1 &: \sigma^2 \neq 1.0 \end{aligned}$$

we compute test statistic

$$q = \frac{(n-1)s^2}{c} = \frac{90.78^2}{1.0} = 5.475$$

and compare with

$$\begin{aligned} C_{R_1} &= F_{\chi^2_{n-1}}(0.025) && \Longrightarrow C_{R_1} = 2.700 \\ C_{R_2} &= F_{\chi^2_{n-1}}(0.975) && \Longrightarrow C_{R_2} = 19.022 \end{aligned}$$

so q is not a suprising observation from a χ^2_{n-1} distribution, and hence we cannot reject H_0.

5.5.5 Two sample tests

It is straightforward to extend the ideas from the previous sections to two sample situations where we wish to compare the distributions underlying two data samples. Typically, we consider sample one, $x_1, ..., x_{n_X}$, from a $N(\mu_X, \sigma_X^2)$ distribution, and sample two, $y_1, ..., y_{n_Y}$, independently from a $N(\mu_Y, \sigma_Y^2)$ distribution, and test the equality of the parameters in the two models. Suppose that the sample mean and sample variance for samples one and two are denoted $(\bar{x}, s_X^2)$ and $(\bar{y}, s_Y^2)$ respectively.

First, consider testing the hypothesis

$$\begin{aligned} H_0 &: \mu_X = \mu_Y \\ H_1 &: \mu_X \neq \mu_Y \end{aligned}$$

when $\sigma_X = \sigma_Y = \sigma$ is known. Now, we have from the sampling distributions theorem we have

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma^2}{n_X}\right) \quad \bar{Y} \sim N\left(\mu_Y, \frac{\sigma^2}{n_Y}\right) \quad \Longrightarrow \bar{X} - \bar{Y} \sim N\left(0, \frac{\sigma^2}{n_X} + \frac{\sigma^2}{n_Y}\right)$$

and hence

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}} \sim N(0,1)$$

giving us a test statistic z defined by

$$z = \frac{\bar{x} - \bar{y}}{\sigma\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}$$

which we can compare with the standard normal distribution; if z is a suprising observation from $N(0,1)$, and lies outside of the critical region, then we can reject H_0. This procedure is the Two Sample Z-Test.

If $\sigma_X = \sigma_Y = \sigma$ is unknown, we parallel the one sample T-test by replacing σ by an estimate in the two sample Z-test. First, we obtain an estimate of σ by "pooling" the two samples; our estimate is the **pooled estimate**, s_P^2, defined by

$$s_P^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}$$

which we then use to form the test statistic t defined by

$$t = \frac{\bar{x} - \bar{y}}{s_P\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}$$

It can be shown that, if H_0 is true then t should be an observation from a Student-t distribution with $n_X + n_Y - 2$ degrees of freedom. Hence we can derive the critical values from the tables of the Student-t distribution.

If $\sigma_X \neq \sigma_Y$, but both parameters are known, we can use a similar approach to the one above to derive test statistic z defined by

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$$

which has an $N(0, 1)$ distribution if H_0 is true.

Clearly, the choice of test depends on whether $\sigma_X = \sigma_Y$ or otherwise; we may test this hypothesis formally; to test

$$\begin{array}{ll} H_0 & : \sigma_X = \sigma_Y \\ H_1 & : \sigma_X \neq \sigma_Y \end{array}$$

we compute the test statistic

$$q = \frac{s_X^2}{s_Y^2}$$

which has a null distribution known as the **Fisher** or F distribution with $(n_X - 1, n_Y - 1)$ degrees of freedom; this distribution can be denoted $F(n_X - 1, n_Y - 1)$, and its quantiles are tabulated. Hence we can look up the 0.025 and 0.975 quantiles of this distribution (the F distribution is not symmetric), and hence define the critical region; informally, if the test statistic q is very small or very large, then it is a suprising observation from the F distribution and hence we reject the hypothesis of equal variances.

5.5.6 Confidence Intervals

The procedures above allow us to test specific hypothesis about the parameters of probability models. We may complement such tests by reporting a **confidence interval**, which is an interval in which we believe the "true" parameter lies with high probability. Essentially, we use the sampling distribution to derive such intervals.

For example, in a one sample Z-test, we saw that

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

that is, that, for critical values $\pm C_R$ in the test at the 5

$$\mathrm{P}[\ -C_R \leq Z \leq C_R\ ] = \mathrm{P}\left[\ -C_R \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq C_R\ \right] = 0.95$$

Now, from tables we have $C_R = 1.96$, so re-arranging this expression we obtain

$$\mathrm{P}\left[\ \bar{X} - 1.96\frac{\sigma}{\sqrt{n}}\ \leq\ \mu\ \leq\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\ \right] = 0.95$$

from which we deduce a **95 % Confidence Interval** for μ based on the sample mean $\bar{x}$ of

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

We can derive other confidence intervals (corresponding to different significance levels in the equivalent tests) by looking up the appropriate values of the critical values. The general approach for construction of confidence interval for generic parameter θ proceeds as follows. From the modelling assumptions, we derive a **pivotal quantity**, that is, a statistic, T_{PQ}, say, (usually the test statstic random variable) that depends on θ, but whose sampling distribution is "parameter-free" (that is, does not depend on θ). We then look up the critical values C_{R_1} and C_{R_2}, such that

$$\mathrm{P}[\ C_{R_1} \leq T_{PQ} \leq C_{R_2}\ ] = 1 - \alpha$$

where α is the significance level of the corresponding test. We then rearrange this expression to the form

$$\mathrm{P}[\ c_1 \leq \theta \leq c_2\ ] = 1 - \alpha$$

where c_1 and c_2 are functions of C_{R_1} and C_{R_2} respectively. Then a $1-\alpha$ % Confidence Interval for θ is $[c_1, c_2]$.

For the tests discussed in previous sections, the calculation of the form of the confidence intervals is straightforward: in each case, C_{R_1} and C_{R_2} are the $\alpha/2$ and $1-\alpha/2$ quantiles of the distribution of the pivotal quantity.

Test	Pivotal Quantity T_{PQ}	Distribution	Parameter	Confidence Interval
ONE SAMPLE Z	$Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	$N(0,1)$	μ	$\bar{x} \pm C_R\ \sigma/\sqrt{n}$
ONE SAMPLE T	$T = \dfrac{\bar{X}-\mu}{s/\sqrt{n}}$	$St(n-1)$	μ	$\bar{x} \pm C_R\ s/\sqrt{n}$
ONE SAMPLE σ	$Q = \dfrac{(n-1)s^2}{\sigma^2}$	χ^2_{n-1}	σ^2	$\left[\dfrac{(n-1)s^2}{C_{R_2}} : \dfrac{(n-1)s^2}{C_{R_1}}\right]$
$\sigma_X = \sigma_Y = \sigma$ known TWO SAMPLE Z	$Z = \dfrac{(\bar{X}-\mu_X)-(\bar{Y}-\mu_Y)}{\sigma\sqrt{\dfrac{1}{n_X}+\dfrac{1}{n_Y}}}$	$N(0,1)$	$\mu_X-\mu_Y$	$(\bar{x}-\bar{y}) \pm C_R\ \sigma\sqrt{\dfrac{1}{n_X}+\dfrac{1}{n_Y}}$
$\sigma_X = \sigma_Y = \sigma$ unknown TWO SAMPLE T	$T = \dfrac{(\bar{X}-\mu_X)-(\bar{Y}-\mu_Y)}{s_P\sqrt{\dfrac{1}{n_X}+\dfrac{1}{n_Y}}}$	$St(n_X+n_Y-2)$	$\mu_X-\mu_Y$	$(\bar{x}-\bar{y}) \pm C_R\ s_P\sqrt{\dfrac{1}{n_X}+\dfrac{1}{n_Y}}$
$\sigma_X \neq \sigma_Y$ known TWO SAMPLE Z	$Z = \dfrac{(\bar{X}-\mu_X)-(\bar{Y}-\mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n_X}+\dfrac{\sigma_Y^2}{n_Y}}}$	$N(0,1)$	$\mu_X-\mu_Y$	$(\bar{x}-\bar{y}) \pm C_R\sqrt{\dfrac{\sigma_X^2}{n_X}+\dfrac{\sigma_Y^2}{n_Y}}$
TWO SAMPLE σ	$Q = \dfrac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2}$	$F(n_X-1, n_Y-1)$	$\dfrac{\sigma_X^2}{\sigma_Y^2}$	$\left[\dfrac{s_X^2}{C_{R_2}s_Y^2} : \dfrac{s_X^2}{C_{R_1}s_Y^2}\right]$