10 Hypothesis Testing

(Reading Exercises: Yates and Goodman – Chapter 13)
(Montgomery and Runger – Sections 9.1.1-9.1.3, 9.1.5, 9.2.1 9.3.1)
A statistical Hypothesis is an assertion or conjecture concerning one or more populations. We take a sample from the population and examine it in order to prove whether a hypothesis is true or untrue.

Definition: Statistical or hypotheses tests are procedures that enable us decide either to reject or to accept claims or assertions about a population.

- Hypothesis testing is a systematic way to test claims or ideas about a population.

Learning outcomes: Students are expected to be able to do the following:

You will be able to

- Define null hypothesis and alternative hypothesis,
- Define Type I error and Type II error
- Identify the steps involved in the hypothesis testing procedure.
- Define level of significance and test statistic.
- Determine critical intervals for known/unknown mean and known/unknown variance.
- Describe the Chi-squared and Student distributions
- Distinguish between one-sided and two-sided hypothesis tests.

10.1 Binary Hypothesis Testing:

In binary hypothesis testing, we are testing two possibilities – reject or accept the hypothesis. There are two possibilities for which the hypothesis is true, and two possibilities for which the hypothesis is untrue.

Consider the two models (a) null hypothesis denoted H_0 (accept hypothesis) and (b) alternate hypothesis denoted H_1 (reject hypothesis). For example, we can partition the space of all possible data sets into two regions – acceptance region and rejection region. In the test, four possible situations can occur:

- H_0 is true but we reject it and accept H_1. This is referred to as Type I error (false rejection or false alarm.
- H_0 is false but we accept it and reject H_1. This is referred to as Type II error (false acceptance or missed detection.
- H_0 is true and we accept it. This is correct decision.
- H_1 is true and we accept it. This is correct decision.

Probability of False Alarm:

This is the probability of Type I error, defined as the conditional probability

$$P_{FA} = P\left[H_1 \text{ is declared true} \mid H_0\right] = \alpha$$

- The probability of committing a Type I error (false alarm), denoted as α , is called the significance level.

Probability of Miss:

This is the probability of Type II error, defined as the conditional probability

$$P_{MISS} = P\left[H_0 \text{ is declared true} \mid H_1\right] = \beta$$

- The probability of committing a Type II error, denoted as β , is called the power of the test.
- We have control of α but not β.

Definition: A plot of P_{MISS} versus P_{FA} is known as the receiver operating curve (ROC)

- It shows the trade-off between probability of miss and probability of false alarm.

Steps in Hypothesis Testing:

Following are the steps for a binary hypothesis test:

1. Collect samples.
2. Formulate the null (H_0) and alternate hypothesis (H_1).
3. Compute the appropriate test statistic.
4. Select a significance level α.
5. Determine the critical value y_c based on the significance level.
6. Establish the acceptance region, $-y_c \le Y \le y_c$ or the the critical region, $Y < -y_c$ and $Y > y_c$
7. Accept H_0 if test statistic is in the acceptance region, otherwise reject if it is in the critical region.
8. Draw engineering conclusions.

Two-sided versus one-sided tests:

In a two-sided test, there are two critical regions for the alternative hypothesis– one in the right tail and the other in the left tail of the probability distribution curve.

In a one-sided test, there is only one critical region for the alternative hypothesis– either in the right tail or in the left tail of the probability distribution curve

Hypothesis Testing of Mean (known variance):

Two sided test: True Mean is unknown and true variance is known:
Here we are testing for the mean since it is unknown. We are given a set of random samples $X = \{x_1, x_2, \ldots, x_n\}$, where the true mean of the population is unknown but the true variance σ_X^2, is known. We want to test the hypothesis that the mean is $\mu_X = \mu_0$.

(1) Formulate the hypotheses:
The hypotheses are briefly stated as follows:

$$\begin{aligned} H_0: &\quad \mu_X = \mu_0 \\ H_1: &\quad \mu_X \neq \mu_0 \quad (\mu_X < \mu_0 \text{ or } \mu_X > \mu_0) \end{aligned}$$

(2) For the test statistic we compute the sample mean $\bar{X}$. Collect n independent samples, $x_1, x_2, \ldots, x_n$ and compute

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The normalized test statistic Y, is given by

$$Y = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Since $\bar{X}$ is Gaussian (central limit theorem) the random variable Y is standard Gaussian. This normalization enables us to use the Q-function table for the test.

(3) Select the confidence level by setting the Type I error probability to some low acceptable value α. From the Q-function table find the critical value y_c, as follows

$$P[-y_c \leq Y \leq y_c] = 1 - \alpha = 1 - 2Q(y_c) \Rightarrow Q(y_c) = 0.5\alpha$$

Look up the value of y_c for which $Q(y_c) = 0.5\alpha$.

(4) Establish the acceptance region for the sample mean as follows:

$$-y_c \le Y = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \le y_c \quad \Rightarrow \quad \mu_0 - \frac{\sigma_X}{\sqrt{n}} y_c \le \bar{X} \le \mu_0 + \frac{\sigma_X}{\sqrt{n}} y_c$$

(5) Make the decision as follows:

- Accept μ_0 (hypothesis H_0) if the value of the sample mean $\bar{X}$ falls inside the acceptance region.
- Reject μ_0 (hypothesis H_1) if the value of the sample mean $\bar{X}$ falls in the critical region.

Example: Suppose a manufacturing line produces resistors that are supposed to be 10Ω. Ten resistors are taken from the production line and measured, with the following results:

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
9.86	9.90	9.93	9.95	9.96	9.97	9.98	10.01	10.02	10.04

Assume each measurement is the actual resistance, $\mu_X = R$, plus a Gaussian measurement error that has mean zero and a variance of 0.1. Test the hypothesis that the resistance is $10\ \Omega$ versus the hypothesis that it is not $10\ \Omega$ at a significance level of 0.05 (95% confidence).

Solution: The hypotheses are

$$H_0: \qquad \mu_X = 10$$
$$H_1: \quad \mu_X \ne 10 \quad (\mu_X < 10 \text{ or } \mu_X > 10)$$

Test statistic: The sample mean is

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} x_i = 9.962$$

Acceptance (or Critical) interval: The acceptance interval for the sample estimate is obtained as

$$-y_c \le Y = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \le y_c \quad \Rightarrow \quad \mu_0 - \frac{\sigma_X}{\sqrt{n}} y_c \le \bar{X} \le \mu_0 + \frac{\sigma_X}{\sqrt{n}} y_c$$

For $\alpha = 0.05$, we find from the Q-function table that $y_c \simeq 1.975$.

$$10 - \frac{\sqrt{0.1}}{\sqrt{10}} y_c \le \overline{X} \le 10 + \frac{\sqrt{0.1}}{\sqrt{10}} y_c \quad \Rightarrow \quad 9.8025 \le \overline{X} \le 10.1975$$

Decision: Since $\overline{X} = 9.962$ falls inside this acceptance interval, we accept the hypothesis that the resistance is 10Ω.

Example: A batch of 100 resistors have an average of 102Ω . Assuming a population standard deviation of 8Ω, test whether the population mean is 100Ω at a significance level $0.95\ (\alpha = 0.05)$.

Solution:

Hypotheses:
$$\begin{aligned} H_0 &: \quad \mu_X = 100 \\ H_1 &: \quad \mu_X \ne 100 \end{aligned}$$

Test statistic: For the test statistic, we compute the sample mean,

$$\overline{X} = \frac{1}{100} \sum_{i=1}^{100} x_i = 102$$

Acceptance region: For $\alpha = 0.05$, $Q(y_c) = 0.5\alpha = 0.025$ and from the Q-function table $y_c \simeq 1.96$

$$100 - \frac{8}{10} \times 1.96 \le \overline{X} \le 100 + \frac{8}{10} \times 1.96 \quad \Rightarrow \quad 98.432 \le \overline{X} \le 101.568.$$

Decision procedure: Since $\overline{X} = 102$ is outside the acceptance region $98.432 \le \overline{X} \le 101.568$, we reject the hypothesis.

Example: We have a production line of resistors that are supposed to be $100\ \Omega$. The mean is unknown and the standard deviation is assumed to be $\sigma_X = 8\ \Omega$. We are given the acceptance region for the sample mean as $98 \le \overline{X} \le 102$ and a sample size of 100 . Find the corresponding significance level or false alarm probability α.

Solution:

Hypotheses:
$$\begin{cases} H_0: & \mu_X = 100 \\ H_1: & \mu_X \ne 100 \ \ (\mu_X < 100 \text{ or } \mu_X > 100) \end{cases}$$

Test statistic: We first compute the sample mean $\overline{X}$,

$$\overline{X} = \frac{1}{100}\sum_{i=1}^{100} x_i$$

We know that the sample mean has a Gaussian distribution with mean μ_X and standard deviation $\sigma_X/\sqrt{n}$. The probability of Type I error is the sum of the areas of the two tails; the critical regions, that is,

$$\begin{aligned}\alpha &= P\left[\overline{X} < 98\right] + P\left[\overline{X} > 102\right] = 1 - P\left[\overline{X} > 98\right] + P\left[\overline{X} > 102\right] \\ &= 1 - Q\left(\frac{98-100}{8/\sqrt{100}}\right) + Q\left(\frac{102-100}{8/\sqrt{100}}\right) = 2Q\left(\frac{2}{8/\sqrt{100}}\right) = 2Q(2.5) = 0.0124\end{aligned}$$

Acceptance region:

$$100 - \frac{8}{10}\times 1.96 \le \overline{X} \le 100 + \frac{8}{10}\times 1.96 \quad \Rightarrow \quad 98.432 \le \overline{X} \le 101.568.$$

Decision procedure: Since $\overline{X} = 102$ is outside the acceptance region, $98.432 \le \overline{X} \le 101.568$, we reject the hypothesis.

One-sided Test of Sample Mean (variance known):

There are two possible one-sided hypothesis testing problems. The first one is

$$\begin{cases} H_0: & \mu_X = \mu_0 \\ H_1: & \mu_X > \mu_0 \end{cases}$$

This is a one-sided test where the acceptance region is on the left side (critical region is on the right side) of the test statistic. For this hypothesis, the acceptance region is

$$\overline{X} \le \mu_0 + \frac{\sigma_X}{\sqrt{n}} y_c$$

The second one is

$$\begin{cases} H_0: & \mu_X = \mu_0 \\ H_1: & \mu_X < \mu_0 \end{cases}$$

This is a one-sided test where the acceptance region is on the right side (critical region is on the left side) of the test statistic.
For the above hypothesis the acceptance region is

$$\bar{X} \geq \mu_0 - \frac{\sigma_X}{\sqrt{n}} \mathrm{y}_c$$

Example: A quality control engineer finds that a sample of 100 light bulbs had an average life-time of 470 hours. Assuming a population standard deviation of 25 hours, test whether the population mean is $\mu_X = 48$ hours versus the alternate hypothesis $\mu_X < 48$ hours at a significance level of $0.95 \Rightarrow \alpha = 0.05$.

Solution:

Hypotheses: $\begin{cases} H_0: & \mu_X = 480 \\ H_1: & \mu_X < 480 \end{cases}$

Test statistic: For the test statistic we first compute the sample mean $\bar{X}$,

$$\bar{X} = \frac{1}{100}\sum_{i=1}^{100} x_i = 470$$

Acceptance region: For $\alpha = 0.05$, $Q(y_c) = \alpha = 0.05$ and from the Q-function table $y_c = 1.65$.

$$\bar{X} \geq 480 - 1.65 \times \frac{25}{10} \quad \Rightarrow \quad \bar{X} \geq 475.875$$

Decision: Since $\bar{X} = 470$ is outside the acceptance region, we reject the hypothesis.

Exercise: A manufacturer of sports equipment has developed a new synthetic fishing line that the company claims has a mean breaking strength of 8 kilograms with a standard deviation of 0.5 kilograms. Suppose a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kilograms. Test the hypothesis that the mean is 8 kilograms versus the alternative that the mean is not 8 kilograms. Use a significance level of $\alpha = 0.01$.

Tests Concerning Sample Mean (Variance unknown):

Hypotheses: $\begin{cases} H_0: & \mu_X = \mu_0 \\ H_1: & \mu_X > \mu_0 \end{cases}$

<u>The test statistic:</u> If the standard deviation is known then the normalized test statistic $\bar{Y}$, is Gaussian,

$$Y = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

However, since the standard deviation is unknown, it must be estimated. The unbiased estimate is the sample variance given by

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2 \quad \Rightarrow \quad S_X = \sqrt{S_X^2}$$

In this case the normalized test statistic is given by

$$T = \frac{\bar{X} - \mu_0}{\mathrm{S}_X/\sqrt{n}}$$

The normalized test statistic T, in this case, satisfies the Student t-distribution with $v = n-1$ degrees of freedom.

- The student t-distribution arises when we are estimating the mean and the variance is unknown.
- If we use n samples to estimate the mean and variance, then the degree of freedom is $v = n-1$
- Similar to the Q-function, the t-distribution is symmetric with respect to zero. Therefore, we may define the significance level as

$$1 - P\left[-t_c \le T \le t_c\right] = \alpha$$

The t-distribution is also tabulated and so the parameter t_c may be obtained from the table in a similar way as the Q-function. The acceptance region is given by

$$\mu_0 - \frac{\mathrm{S}_X}{\sqrt{n}}t_c \le \bar{X} \le \mu_0 + \frac{\mathrm{S}_X}{\sqrt{n}}t_c$$

For a given significance level α, we look up the value for t_c and determine the acceptance region. We accept the hypothesis if $\bar{X}$ falls inside the acceptance region, otherwise, we reject the hypothesis.

Example: Test the hypothesis that the average content of containers of a particular lubricant is 10 liters if the contents of a random sample of 10 containers are $10.2,\ 9.7,\ 10.1,\ 10.3,\ 10.1,\ 9.8,\ 9.9,\ 10.4,\ 10.3,\ 9.8$ liters. Use a confidence level of $0.99\ (\alpha = 0.01)$ and assume that the distribution of contents is Gaussian.

Solution: $\alpha = 0.01$ **and** $v = 9$. From t-tables, $t_c = 3.25$

$$H_0: \quad \mu_X = 10,$$
$$H_1: \quad \mu_X \neq 10.$$

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 10.06$$

$$S_X = \sqrt{\frac{1}{9}\sum_{i=1}^{10}\left(X_i - 10.06\right)^2} = 0.246$$

$$10 - \frac{S_X}{\sqrt{n}}t_c \le \bar{X} \le 10 + \frac{S_X}{\sqrt{n}}t_c$$

$$10 - \frac{0.246}{\sqrt{10}}\times 3.25 \le \bar{X} \le 10 + \frac{0.246}{\sqrt{10}}\times 3.25 \ \Rightarrow\ 9.75 \le \bar{X} \le 10.25$$

Since the value of $\bar{X}$ falls within the critical interval, we accept the hypothesis that $\mu_X = 10 \text{ liters}$.