9 Hypothesis Tests

1. Inference requires a prior "tentative assumption".

   Definition 9.1. A *hypothesis* is a statement about a population parameter.

2. The key point is that hypothesis makes a statement about the population and the goal of such tests is to decide based on a sample from the population, which of the two complementary hypothesis is true.

   Definition 9.2. The two complementary hypothesis in a hypothesis testing problem are called the *null hypothesis* and the *alternative hypothesis*.

3. If θ denotes the population parameter, the general format of the null and alternative hypothesis is $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_0^c$, where Θ_0 is some subset of parameter space and H_0 and H_1 denote the null and the alternative hypothesis respectively.

4. *Example*: If we are interested in the average probability, denoted by θ of getting infected by the coronavirus after taking a Pfizer vaccine, we might be interested in testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$.

5. In a hypothesis testing problem, after observing the sample the experimenter must decide either to accept H_0 as true or to reject H_0 as false and decide H_1 is true.

   Definition 9.3. A hypothesis testing procedure or a *hypothesis test* is a rule that specified:

   (a) For which sample values the decision is made to accept H_0 as true.

   (b) For which sample values H_0 is rejected and H_1 is accepted as true.

6. The subset of the sample space for which H_0 will be rejected is called the *rejection region* or the critical region. Its complement is called the acceptance region.

7. A Hypothesis test is specified in terms of a *test statistic* $W(X_1, X_2, \ldots, X_n)$ a function of the sample.

8. A test might specify that the null H_0 is to be rejected if $\bar{X}$, the sample mean, is greater than 10. In this case W($\mathbf{X}$) $= \bar{X}$ is the test statistic and the rejection region is $\{(x_1, x_2, x_3, \ldots, x_n) : \bar{x} > 10\}$

9.1 Developing Null and Alternative Hypotheses

1. The *context of the situation* is very important in determining how the hypotheses should be stated.

2. *Example*: Honda Jazz Mileage

   (a) Consider the case of Honda Jazz which has a mileage of 24 miles per gallon while driving in the city.

   (b) The car manufacturer hires Mr. Yu to test whether their new i-VTEC Jazz engine has a better mileage than the comparable older Jazz cars which gave a mileage of 24 mpg.

   (c) Yu asks for a few new versions of Jazz to make a test hypothesis and a couple of R2D2s to drive the new cars.

   (d) The sample mean miles per gallon for these automobiles will be computed and used in a hypothesis test to determine if it can be concluded that the new system provides more than 24 miles per gallon.

(e) The *research hypothesis* (the hypothesis you think is true) $\mu > 24$, when μ is the population mean miles per gallon becomes the alternative hypothesis, when Yu is trying to gather evidence on the research hypothesis.

(f) Alternatively, Yu could confront this with a tentative assumption which he has to disprove - $\mu \leq 24$. The final hypothesis tests in this case are:

$$H_0 : \mu \leq 24$$
$$H_1 : \mu > 24$$

(g) Finally, iff the sample results lead to the conclusion to reject H_0, the inference can be made that $H_1 : \mu > 24$ is true. The researchers have the statistical support to state that the new fuel injection system increases the mean number of miles per gallon.

3. Null first: consider an example where we have no research hypothesis, but instead have a prior about the population parameter. In such instances, usually null hypothesis represents the prior that you have and we come up with a complimentary alternative.

4. *Example* of Food Labels: consider a label of a Cola which has 67.6 fluid ounces (approximately 2L). We think that the label is accurate if the minimum contained in the bottle is *at least* 67.6.

Figure 2: Learn more about rules of a food label here

(a) We begin with the assumption that label is correct and state the null as $H_0 : \mu \geq 67.6 fl.ou.$

(b) A rejection of the null indicates a consistent under filling of the bottles and a violation represented by the alternative as $H_1 : \mu < 67.6$.

(c) The government agency – the Food Safety and Standards Authority of India under the Ministry of Health and Family Welfare – operates under the Food Safety and Standards Regulation, 2011 for India needs to check these claims occasionally.

(d) FSAAI collects samples, compute the sample mean filling weight, and use the sample results to test the preceding hypotheses.

(e) If results indicate a rejection of H_0, they must take legal action against the companies.

Task: Check whether the FSAAI makes this data public and how many times does it undertake such checks?

5. *Example* of under and over filling: from the point of view of manufacture is bad. Underfilling could lead to litigation and overfilling causes loss of profits.

   (a) Company begins with the assumption that they are correctly filling the bottles and a null hypothesis of $H_0 : \mu = 0$.

   (b) The alternative which is cautious of erring on both sides is $H_1 : \mu \neq 0$.

   (c) Company begins quality control by collecting a sample from its plant and computing the average filling per bottle.

   (d) If the sample leads to rejection of H_0 then we know that its quality control has indicated an error in the assembly line which needs urgent correction.

6. *Context of situation* and the *point of view of the investigator* is important in formulating the hypothesis.

7. In this chapter we learn about hypothesis on population mean and population proportions

   We distinguish between the types of Hypothesis

$$\begin{array}{lll} H_0 : \mu \geq \mu_0 & H_0 : \mu \leq \mu_0 & H_0 : \mu = \mu_0 \\ H_a : \mu < \mu_0 & H_a : \mu > \mu_0 & H_a : \mu \neq \mu_0 \end{array}$$

   First two are called one tailed tests, Last is called the two tailed test.

   Note that as the preceding forms show, the equality part of the expression (either $\geq, \leq$ or $=$) always appears in the null hypothesis

9.2 Type I and Type II Errors

1. We need to test the hypothesis, once you have formulated it. Unfortunately, the correct conclusions are not always possible. The two types of errors have been given the non-mnemonic names, Type-I and Type-II error.

2. Type I error: is we reject H_0 when its actually true (false negative). If $\theta \in \Theta_0$, but the hypothesis test incorrectly decided to reject H_0, then a Type-I error has been made,

3. Type II error: $\theta \in \Theta_0^c$ but the test decides to accept H_0, a Type-II error has been made i.e. a false positive.

TABLE 9.1 Errors and Correct Conclusions in Hypothesis Testing

		Population Condition	
		H_0 True	H_a True
Conclusion	Accept H_0	Correct Conclusion	Type II Error
	Reject H_0	Type I Error	Correct Conclusion

4. *Example*: Honda Jazz Mileage: recall the hypothesis made by Mr. Yu for testing the mileage of Honda jazz with i-Vetc engines.

$$H_0 : \mu \leq 24$$
$$H_1 : \mu > 24$$

In this case, Mr. Yu would ideally like to side with alternative hypothesis which is the research hypothesis in this case.

   (a) Type-I error is when Mr. Yu rejects H_0 and tells customers of the new car that its i-Vtech engine has greater mileage than its previous versions, when in fact the new engine mileage is the same or worse than the old Jazz cars.

   Perhaps the R2D2s messed up.

   (b) Type-II error is when Mr.Yu claims that the new i-Vtech engines are no better than the old ones, when in fact they are significantly better. In such case H_1 is true but you err by siding with H_0

5. Level of significance: The probability of rejecting $H_0 : \mu \leq 24$, *when it is true (as an equality* i.e. $\mu = 24$) is the "level of significance" in our Jazz example. Level of significance is denoted by α and common choices are 0.05 (5%) and 0.01 (1%).

   Definition 9.4. Level of Significance Level of significance is the probability of making a Type 1 error when the null hypothesis is true as an equality

   Hypothesis tests which control for only Type-I error are called 'significance tests'.

6. Semantics: Often tests do not control for Type-II error and hence our decision does not take into consideration all possible forms of errors. Statisticians usually recommend that we use the statement "do not reject H_0" instead of "accept H_0". Using the statement "do not reject H_0" carries the recommendation to withhold both judgment and action.