Ch2 - Testing

ECO 204
Statistics For Business and Economics - II

Shaikh Tanvir Hossain

East **W**est **U**niversity, Dhaka
Last updated: March 17, 2025

Outline

Outline

Testing : Key Ideas

Testing : Key Ideas

What is Testing, Errors and Different Types of Tests

Testing, Errors and Different Types of Tests

- In this chapter we will learn a new technique in inferential statistic, known as *Hypothesis Testing*, or in short often we just say - *Testing*.
- Testing problem is slightly different than estimation, here
  - First, we start with *two competing* hypotheses about the unknown population parameter, one is called *"Null Hypothesis"* and the other one is *"Alternative Hypothesis"*. We then use the sample data to reject or accept the Null Hypothesis.

.

- *Question is* -

  What is a hypothesis?

  Ans: It's simply a conjecture about the population parameter.
- Let's see an example.

Testing, Errors and Different Types of Tests

- Continuing from the problem we introduced in the last chapter, suppose we have an information that the unknown population mean income μ is less than $24,000$, now we would like to test whether $\mu < 24,000$, is true, to do this we can form two hypotheses as follows,

$$H_0 : \mu \geq 24,000 \quad \textit{Null Hypothesis}$$
$$H_a : \mu < 24,000 \quad \textit{Alternative Hypothesis}$$

- Now after we are done with constructing the hypotheses, we use a random sample (or data) to either *reject the Null* or *accept the Null*. Acceptance is sometimes written as *fail to reject*, however in this chapter we will avoid these philosophical issues!

- Two important points,

  i) What we believe is written in the alternative (this is often the convention, however not necessary!)

  ii) Notice! Everything is happening around Null. So we are either rejecting the Null or accepting the Null (Why? We will come back to the answer later!)

- Since in practice we never know where is μ, so in this setting we can make two types of errors, which are known as *Type I Error* and *Type II Error*, and also there are two scenarios where we are correct, let's see this now

Testing, Errors and Different Types of Tests

		Population Reality	
		$\boldsymbol{H}_0$ True	$\boldsymbol{H}_a$ True
Conclusion	Accept $\boldsymbol{H_0}$	Correct Conclusion	Type II Error
	Reject $\boldsymbol{H}_0$	Type I Error	Correct Conclusion

- Before interpreting the table, first of all, always remember

  we do not know what is μ

- Now the table says, if the hypothesized Null is actually true and after the testing we accept Null, then there is no error, but if the hypothesized Null is true and we reject the Null then we will make an error and the error is called Type-I error....Can you interpret other cells of the table?
- Ideally we want to construct a test that minimizes both of these errors, but actually for a fixed sample size, this is impossible. So the idea is we fix the Type-I error and look for a test which minimizes the Type-II error. We won't go to more theoretical details here... in our testing procedure we will always *fix Type-I error* and our procedure will minimize Type-II error...
- Let's replace the hypothesized information $24,000$ with μ_0, essentially this is the value of the unknown parameter where we are dividing the parameter set. In this case we will now think about following three formations of testing

Two Tail test

$$H_0 : \mu = \mu_0$$

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

Upper Tail test

$$H_{0:} : \mu \leq \mu_0$$

$$H_a : \mu > \mu_0$$

Lower Tail test

$$H_{0:} : \mu \geq \mu_0$$

$$H_a : \mu < \mu_0$$

- Together the last two tests are called - *One tail tests*. The word tail is coming from the tail of the Normal distribution...but you will understand later why this naming...

Testing : Key Ideas

Test Statistic and Sampling Distribution in Testing

Test Statistic and Sampling Distribution in Testing

- In the hypothesis testing again we will use the *sampling distribution of $\overline{X}$*. Recall before we called $\overline{X}$ an estimator of μ.
- In the Hypothesis testing we won't call this an estimator, rather we will call it a *Test Statistic*. In general often a *Test Statistic* is same or very similar to a point estimator, but its a *Statistic*, which is simply *a function of the random sample*
- When a *Statistic* is used for estimation we call it an *Estimator*. Similarly, when a *Statistic* is used for *Testing* we call it a *Test Statistic*. These are just some naming conventions that you need to know.

Test Statistic and Sampling Distribution in Testing

- For example $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a function of random sample, so this is a statistic. When we use it for point estimation, we call this point estimator, but if we use it for hypothesis testing we call it a *Test Statistic*.
- Before we said the distribution of the estimator is called *sampling distribution*, actually more generally the probability distribution of a *Statistic* is called *Sampling Distribution*. So here the distribution of $\overline{X}$ is a sampling distribution (old stuff!)
- Again we will use the three old results (below we always assume i.i.d.)

Test Statistic and Sampling Distribution in Testing

- **Result 1:** If the *population data is normal* with μ and σ^2, then $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, we can use this if *we know σ*. In this case we often use Z statistic, where

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ and } Z \sim \mathcal{N}(0, 1)$$

- **Result 2:** If the *population data is normal*, but *we don't know σ*, then we use T statistic, where

$$T = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \text{ and } T \sim t_{n-1}$$

  where S is the sample standard deviation, Note that in these cases,

$$\mathbb{E}(\overline{X}) = \mu \text{ and } \mathbb{V}(\overline{X}) = \frac{\sigma^2}{n}$$

  however in the second case we cannot calculate the variance, hence we use an estimator of the variance, which is S^2/n, and we can write this as,

$$\widehat{\mathbb{V}}(\overline{X}) = \frac{S^2}{n}$$

Test Statistic and Sampling Distribution in Testing

- **Result 3:** If the *population data is not normal* we cannot use *exact distributions* like above, rather we need to *use an approximation following Central Limit Theorem (CLT)*, in that case, we use the statistic

$$Z = \frac{\overline{X} - \mathbb{E}\left(\overline{X}\right)}{\sqrt{\widehat{\mathbb{V}}(\overline{X})}} \text{ when } n \text{ is large,} \qquad \text{so } Z \overset{approx}{\sim} \mathcal{N}(0,1)$$

where $\widehat{\mathbb{V}}(\overline{X})$ is the estimator for the variance of $\overline{X}$. Important is Z follows $\mathcal{N}(0,1)$ when the sample size n becomes very large, this is called *Asymptotic Normality* of the sample mean.

Test Statistic and Sampling Distribution in Testing

CLT in Bernoulli case:
Note it's an approximation and the data may or not be normal. For example, if the population distribution is Bernoulli (i.e., 0 or 1 data, think about Gender), then target parameter is

$$\mathbb{E}(X) = p \text{ where } \mathbb{P}(X = 1) = p$$

In practice we never know p, so we can propose the following estimator for p

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \text{ in this case it is } \textit{sample proportion}$$

It is possible to show that (we skip the proof but it's very easy under i.i.d. assumption)

$$\mathbb{E}(\overline{X}) = p \text{ and } \mathbb{V}(\overline{X}) = \frac{p(1-p)}{n}$$

since we don't know p, we use an estimator of the variance $\mathbb{V}(\overline{X})$, which is

$$\widehat{\mathbb{V}}(\overline{X}) = \frac{\overline{X}(1-\overline{X})}{n}$$

this gives the following Z

$$Z = \frac{\overline{X} - p}{\sqrt{\widehat{\mathbb{V}}(\overline{X})}} \text{ and } Z \overset{approx}{\sim} \mathcal{N}(0,1) \text{ when } n \text{ is large}$$

Test Statistic and Sampling Distribution in Testing

so in this case, in large sample the approximate sampling distribution of the sample proportion is normal with the estimated standard error

$$\sqrt{\widehat{\mathbb{V}}(\overline{X})} = \sqrt{\frac{\overline{X}(1-\overline{X})}{n}}$$

if you are curious about the exact distribution of the sample proportion in finite samples, it is $\mathrm{Bin}(n, p)$ with scaling $\frac{1}{n}$, this means the values will be $0, 1/n, 2/n, \ldots, 1$

Z Test (σ known)

The following example is slightly modified from Anderson et al. (2020)).

The Bangladesh Golf Federation (BGF) establishes rules that manufactures of golf equipment must meet if their products are to be acceptable for use in BGF events. Company ABC uses a high-technology manufacturing process to produce golf balls with a mean driving distance of 295 yards. This is what it advertises to sell its balls.

Sometimes, however, the process gets out of adjustment and produces golf balls with a mean driving distance different from 295 yards. When the mean distance falls below 295 yards, the company worries about losing sales because the golf balls do not provide as much distance as advertised. When the mean distance goes above than 295 yards, ABC also worries because golf balls may be rejected by the BGF for exceeding the overall distance standard concerning carry and roll.

ABC's quality control department takes time to time a sample of 50 golf balls to monitor the manufacturing process. It calculated the sample mean of 50 golf balls and found it is to be 297.6.

For this sample, now the department wants to do hypothesis testing to determine whether the process has fallen out of adjustment. Develop the null and alternative hypotheses and do the testing at 5% level of significance, and also do the test, suppose somehow the quality control team knows that population standard deviation of all golf balls is $\sigma = 12$.

- How do we develop the Null and Alternative Hypothesis? Note that, both below and above of 295 is problematic for the company. So the proper formation is, (notice here $\mu_0 = 295$)

$$H_0 : \mu = 295$$
$$H_a : \mu \neq 295$$

- From the story, we know
  - $n = 50$
  - $\bar{x} = 297.6$
  - $\alpha = 0.05$
  - $\mu_0 = 295$
  - $\sigma = 12$
- What does the company want? It's clear that the company would like to accept the Null. This is because if the company rejects the Null then it's costly, why? maybe because it has to change its entire production process.
- We will directly use standard normal distribution (although not necessary in this case), which we denoted with $\mathcal{N}(0, 1)$ to do the testing of the ABC company, here are the steps

Algorithm 1: Two Tail Test : σ known

Input: $\overline{x}$, σ, n, μ_0, α

Output: Reject H_0 or Accept H_0

Formulate the hypotheses H_0 and H_a properly such that

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

Calculate the value of the Z Statistic which is z_{calc} using the formula

$$z_{\text{calc}} = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

Calculate the critical values $z_{\alpha/2}$ and $z_{1-\alpha/2}$ using $\mathcal{N}(0, 1)$

if $z_{calc} \geq z_{1-\frac{\alpha}{2}}$ *or* $z_{calc} \leq z_{\frac{\alpha}{2}}$ **then**

  Reject H_0

else

  Accept H_0

Recall here the critical values are such that,

$$\mathbb{P}(Z \leq z_{\frac{\alpha}{2}}) = \alpha/2 \quad \text{and} \quad \mathbb{P}(Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \alpha/2$$

here is the picture of critical values...

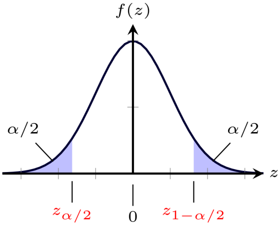

Figure 1: Density Function of Standard Normal Distribution with Two Tail Critical Values

The R code is also very simple, here it is

R code - sigma known, two-tail (using standard normal)

```
# First give the data and calculate zcalc
n <- 50
xbar <- 297.6
alpha <- 0.05
mu0 <- 295
sigma <- 12

zcalc <- (xbar - mu0)/(sigma/sqrt(n))

# check the value
zcalc
# [1] 1.532065

# (alpha/2) quantile of the standard normal
qnorm(alpha/2)
# [1] -1.959964


# (1 - alpha/2) quantile of the standard normal
qnorm(1 - alpha/2)
# [1] 1.959964
```

- In our problem, $z_{\text{calc}} = 1.532$
- Also
  - $z_{\alpha/2} = z_{.025} = -1.96$, this is the .025th quantile or 2.5th percentile of the standard normal distribution.
  - $z_{1-\alpha/2} = z_{.975} = 1.96$, this is the .975th quantile or 97.5th percentile of the standard normal distribution.
- So we can see that

$$-1.96 < \color{red}{1.532} < 1.96$$

or

$$z_{\alpha/2} < z_{\text{calc}} < z_{1-\alpha/2}$$

- So this means our transformed sample mean 1.532 does not fall in the rejection region, so *we accept the Null*....bottomline the ABC company is happy
- Now we can adjust the problem slightly for example,

Problem Changed... (modified for upper-tail test)

ABC only checks whether the average distance is *at max* 295, but worries if the average distance exceeds 295.

- This happens when the company is worried about the golf balls being rejected by the BGF for exceeding the overall distance standard concerning carry and roll. In this case, the company is worried about the average distance exceeding 295 yards. So the proper formation of the hypotheses is, (notice here $\mu_0 = 295$)

$$H_0 : \mu \leq 295$$
$$H_a : \mu > 295$$

- Note the word upper is coming from the alternative hypothesis.
- Similarly, we can think about

Problem Changed... (modified for lower-tail test)

ABC only checks whether the average distance is *at least* 295, but worries if the average distance is below 295.

- This happens when the company is worried about losing sales because the golf balls do not provide as much distance as advertised. In this case, the company is worried about the average distance falling below 295 yards. So the proper formation of the hypotheses is, (notice here $\mu_0 = 295$)

$$H_0 : \mu \geq 295$$
$$H_a : \mu < 295$$

- This is called the lower tail test,
- The steps are similar,

Different z-Tests

First Example and Steps

Algorithm 2: Upper Tail Test : σ known

Input: $\overline{x}$, σ, n, μ_0, α
Output: Reject H_0 or Accept H_0

Formulate the hypotheses H_0 and H_a properly such that

$$H_0 : \mu \leq \mu_0$$
$$H_a : \mu > \mu_0$$

Calculate the value of the Z Statistic which is z_{calc} using the formula

$$z_{\text{calc}} = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$$

Calculate the critical values $z_{\alpha/2}$ and $z_{1-\alpha/2}$ using $\mathcal{N}(0, 1)$
if $z_{calc} \geq z_{1-\alpha}$ **then**
    Reject H_0
else
    Accept H_0

Different z-Tests

First Example and Steps

Algorithm 3: Lower Tail Test : σ known

Input: $\overline{x}$, σ, n, μ_0, α
Output: Reject H_0 or Accept H_0

Formulate the hypotheses H_0 and H_a properly such that

$$H_0 : \mu \geq \mu_0$$
$$H_a : \mu < \mu_0$$

Calculate the value of the Z Statistic which is z_{calc} using the formula

$$z_{\text{calc}} = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$$

Calculate the critical values $z_{\alpha/2}$ and $z_{1-\alpha/2}$ using $\mathcal{N}(0, 1)$
if $z_{calc} \leq z_\alpha$ **then**
    Reject H_0
else
    Accept H_0

- The codes are similar, and you should be able to adjust it... and here the error regions for the upper and lower tail tests,

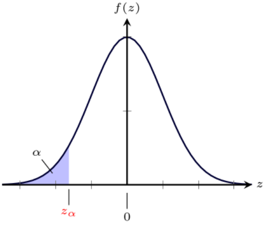

Figure 2: Density Function of Standard Normal Distribution with Lower-Tail Critical Values (Left) and Upper-Tail Critical Values (Right)

t-Tests (σ un-known)

Basics of t-Tests

- Now we will learn t-test where we don't assume that we know the population standard deviation σ, rather we estimate it using s it from the sample data. Recall the formula for s, which is

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

- The whole process will be similar to z-test however we will calculate the value of the T-statistic which is

$$t_{\text{calc}} = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

- And then compare this with the critical t values, for example $t_{1-\alpha/2}$, $t_{\alpha/2}$, t_α and $t_{1-\alpha}$, and then decide whether to reject the Null or not. Note that these values are coming from the t-distribution, which is denoted as t_{n-1}, where $n-1$ is the degrees of freedom (in short df)
- If you want to calculate the critical values from the t_{n-1} using R then, the function is `qt`, in this case,

$$t_{\alpha/2} = \texttt{qt}(\alpha/2, n-1)$$
$$t_{1-\alpha/2} = \texttt{qt}(1-\alpha/2, n-1) \text{ and so on...}$$

Basics of t-Tests

- You should remember that we are still on the assumption that the data is coming from the normal distribution, however unlike the Z test, in this case we don't assume that we know the population standard deviation σ, and this is why according to the Theorem 1.2, we are using the t-distribution. Recall the random quantity in this case is,

$$T = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$$

  and

$$T \sim t_{n-1}$$

- Finally, again, the whole testing procedure for two-tail and one-tail tests are similar, so I won't repeat them in the slides here.

What is the Intuition Behind Hypothesis Testing?

Intuition Behind Hypothesis Testing

- Recall whenever we are doing testing, we are always rejecting or accepting H_0 (or Null), in the math we are also using μ_0 which is coming from when unknown $\mu = \mu_0$
- In fact in testing we are actually using *sampling distribution under the Null*, and we are checking whether the sample mean falls in the rejection region or not. The α is the rejection region which is also an area when we are looking at the Null distribution (and this is type-1 error)
- The idea is when our sample mean is coming from the extreme part, chances are that it's not coming from the Null distribution, and this is why we reject the Null. Similarly when our sample mean is coming from the middle part, chances are that it's coming from the Null distribution, and this is why we accept the Null, following picture will be useful,

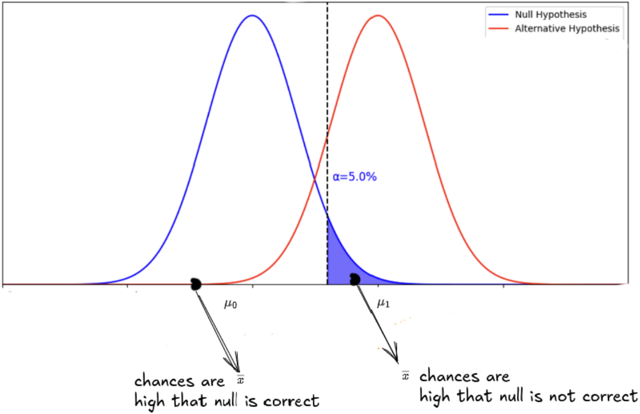

Figure 3: The Sampling Distribution of $\overline{X} \sim \mathcal{N}(\mu_0, \sigma^2/n)$ on the left (this is under H_0), and $\overline{X} \sim \mathcal{N}(\mu_1, \sigma^2/n)$ on the right (this is one possible alternative)

Intuition Behind Hypothesis Testing

- The reason we are looking at the null since this is distribution where we can check, in H_1 we have too many possibilities...

Proportion Testing

Testing for Population Proportion p

- The problem of Population Proportion is similar, but here we have a Bernoulli random variable,
- Suppose we would like to test that the population proportion of the female students at EWU is 64%,
- In this case we can think about a Bernoulli random variable X that represents Gender of a student, where $X = 1$ means female and $X = 0$ means male.
- In this case the population mean, is

$$\mathbb{E}(X) = \mathbb{P}(X = 1) = p$$

  where p is the population proportion of the female students at EWU.
- And we also know that the variance of the Bernoulli random variable is

$$\mathbb{V}(X) = p(1 - p)$$

- In this case we also calculate the sample mean $\overline{X}$, which is the sample

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- However the sample mean in this case is the sample proportion of the female students (do you see why?)

Testing for Population Proportion p

- Again assuming we have a sample of n students, and we have an iid sample, the theory tells us that (Look at Theorem 1.1)

$$\mathbb{E}(\overline{X}) = p \text{ and } \mathbb{V}(\overline{X}) = \frac{p(1-p)}{n}$$

- Question is what is the sampling distribution of $\bar{X}$, here we will use Theorem 1.3, which is a large sample result, which says, if we construct the Z Statistic like following,

$$Z = \frac{\overline{X} - \mathbb{E}(\overline{X})}{\sqrt{\mathbb{V}(\overline{X})}} = \frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

- Then if the sample size is large, or $n \to \infty$, we have

$$Z \overset{approx}{\sim} \mathcal{N}(0, 1)$$

- This means if the sample size is large, then in the large sample we can again use Normal distribution,

Testing for Population Proportion p

Here is an applied problem for proportion testing, this is question 38 from Chapter 9.5

A study by Consumer Reports showed that 64% of supermarket shoppers believe supermarket brands of ketchup to be as good as national name brands. To investigate whether this result applies to its own product, the manufacturer of a national name-brand ketchup asked a sample of shoppers whether they believed that supermarket ketchup was as good as the national brand ketchup.

a. Formulate the hypotheses that could be used to determine whether the percentage of supermarket shoppers who believe that the supermarket ketchup was as good as the national brand ketchup differed from 64%.

b. If a sample of 100 shoppers showed 52 stating that the supermarket brand was as good as the national brand, then at $\alpha = .05$ should the national brand ketchup manufacturer be pleased with this conclusion?

- The methods we used to do hypothesis testing (both for z-test and t-test) is known as *critical value approach*.
- There is another approach of doing the test, which is called ***p** value approach*.
- It's important that both way of doing the tests will give you same answer, but p value approach has some advantages.
- *Question: what is a **p**-value?* In simple words, ***p**-value is the lowest probability at which H_0 can be rejected*. Or we can say it is the *smallest significance level at which we can reject the Null*.
- For example, if we are testing a given hypothesis with $\alpha = 0.05$ and we calculated the p value equal to 0.03, then in the $p-$value approach we will compare α with p-value and come to a conclusion.

- The calculation of p-value is quite easy.
- For *z-test* (or when we know σ), if we have calculated z_{calc}, then we need to find following probabilities and these are the p values for a z-test.
  - 1. p value for the both upper and lower tail test: $\mathbb{P}(Z_n > |z_{calc}|)$
  - 2. p value for a two-tail test: $2 \times \mathbb{P}(Z_n > |z_{calc}|)$.
- So you should think p value is like a probability and it should always be between 0 and 1.
- The rejection rule using p-values is always same that is

$$\text{reject the Nulll } H_0 \text{ if } p - \text{value } < \alpha$$

- In the p-value approach we are comparing probabilities vs. probabilities, where in critical value approach, we are comparing z values vs. z values.
- For the *t-test* the calculation is similar
  - 1. p value for both the upper and lower tail test: $\mathbb{P}(T_n > |t_{calc}|)$
  - 2. p value for a two-tail test: $2 \times \mathbb{P}(T_n > |t_{calc}|)$
- The rejection rule using p-values is always same that is

$$\text{reject the Nulll } H_0 \text{ if } p - \text{value } < \alpha$$

- There is one clear benefit of the p value approach over the critical value approach, that is if we are asked to do the same test for different α, for example we are doing the same test for $\alpha = .10$, $\alpha = .05$ and $\alpha = .01$, in the p-value approach we can compute the p-value only once and the can compare it different α values. Only one calculation is enough. This is very convenient.
- In the critical value approach, every time we change α, we need to calculate the critical values again to do the test. For example if we change α, then we need to calculate z_α or $z_{1-\alpha}$ or $z_{\alpha/2}$ or $z_{1=\alpha/2}$ again to do the test. This is very cumbersome....

References

Abraham, B. and Ledolter, J. (2006), *Introduction to Regression Modeling*, Duxbury applied series, Thomson Brooks/Cole, Belmont, CA.

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., Cochran, J. J., Fry, M. J. and Ohlmann, J. W. (2020), *Statistics for Business & Economics*, 14th edn, Cengage, Boston, MA.

Bertsekas, D. and Tsitsiklis, J. N. (2008), *Introduction to probability*, 2nd edn, Athena Scientific.

Blitzstein, J. K. and Hwang, J. (2015), *Introduction to Probability*.

Casella, G. and Berger, R. L. (2002), *Statistical Inference*, 2nd edn, Thomson Learning, Australia ; Pacific Grove, CA.

DeGroot, M. H. and Schervish, M. J. (2012), *Probability and Statistics*, 4th edn, Addison-Wesley, Boston.

Hansen, B. (2022), *Econometrics*, Princeton University Press, Princeton.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2023), *An introduction to statistical learning*, Vol. 112, Springer.

Newbold, P., Carlson, W. L. and Thorne, B. M. (2020), *Statistics for Business and Economics*, 9th, global edn, Pearson, Harlow, England.

Pishro-Nik, H. (2016), *Introduction to probability, statistics, and random processes*.

Ramachandran, K. M. and Tsokos, C. P. (2020), *Mathematical Statistics with Applications in R*, 3rd edn, Elsevier, Philadelphia.

References

Rice, J. A. (2007), *Mathematical Statistics and Data Analysis*, Duxbury advanced series, 3rd edn, Thomson/Brooks/Cole, Belmont, CA.

Wasserman, L. (2013), *All of statistics: a concise course in statistical inference*, Springer Science & Business Media.

Wooldridge, J. M. (2009), *Introductory Econometrics: A Modern Approach*, 4th edn, South Western, Cengage Learning, Mason, OH.