# Maths & Stats Pre-Sessional Tutorial

SOLUTIONS to exercise for topic 3: Estimation and Hypothesis Testing

---

**Exercise 1**

a) $E(3X + 2Y) = 3 * E(X) + 2 * E(Y) = 3(0.09) + 2(0.12) = 0.51$

b) $Var(3X + 2Y) = 3^2 Var(X) + 2^2 Var(Y) + 2 * 3 * 2Cov(X,Y) =$
$3^2 Var(X) + 2^2 Var(Y) + 2 * 3 * 2Corr(X,Y)\sqrt{Var(X)}\sqrt{Var(Y)} =$
$9 * (0.2)^2 + 4 * (0.27)^2 + 2 * 3 * 2 * 0.6 * 0.20 * 0.27 = 1.04$
The standard deviation is $\sqrt{1.04} = 1.02$

**Exercise 2**

(a) Mean and variance of the sampling distribution for the sample mean are the following:

$$\mu_{\bar{x}} = \mu = 100$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{900}{30} = 30$$

(b) Since for $n \geq 25$ the Central Limit theorem applies, we have that:

$$z_{\bar{x}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

approaches the standard normal distribution. We can therefore find the $P(\bar{x} > 109)$ as follows:

$$z_{\bar{x}} = \frac{109 - 100}{\sqrt{30}} = 1.64, \quad P(\bar{x} > 109) = 1 - \Phi(1.64) = 1 - 0.9495 = 0.0505$$

(c) In order to find $P(96 \leq \bar{x} \leq 110)$ we first find the probabilities of the two events as follows:

$$z_{\bar{x}} = \frac{96 - 100}{\sqrt{30}} = -0.73 \quad \& \quad z_{\bar{x}} = \frac{110 - 100}{\sqrt{30}} = 1.83$$

and finally:

$$P(96 \leq \bar{x} \leq 110) = \Phi(1.83) - \Phi(-0.73) = 0.96638 - 0.2327 = 0.7337$$

(d) $P(\bar{x} \leq 107) = \Phi(z_{\bar{x}}) = \Phi(\frac{107-100}{\sqrt{30}}) = \Phi(1.28) = 0.89973.$

## Exercise 3

(a)

$$P(\bar{x} > 210,000) = P\Big(z > \frac{210,000 - 215,000}{25,000/\sqrt{100}}\Big) = P(z > -2) = 0.9772$$

(b)

$$\begin{aligned} P(213,000 < \bar{x} < 217,000) &= P\Big(\frac{213,000 - 215,000}{25,000/\sqrt{100}} < z < \frac{217,000 - 215,000}{25,000/\sqrt{100}}\Big) \\ &= P(-0.8 < z < 0.8) \\ &= \Phi(0.8) - \Phi(-0.8) \\ &= 0.78814 - 0.21186 \\ &= 0.5763 \end{aligned}$$

(c)

$$\begin{aligned} P(214,000 < \bar{x} < 216,000) &= P\Big(\frac{214,000 - 215,000}{25,000/\sqrt{100}} < z < \frac{216,000 - 215,000}{25,000/\sqrt{100}}\Big) \\ &= P(-0.4 < z < 0.4) \\ &= \Phi(0.4) - \Phi(-0.4) \\ &= 0.65542 - 0.34458 \\ &= 0.3108 \end{aligned}$$

(d) The sample mean selling price is most likely to lie in the range $[214,000; 216,000]$ since it is centred around the given population mean.

(e) The results would still be valid because the central limit theorem applies with the sample size being larger than 30.

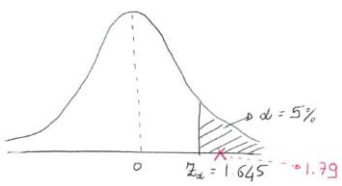
## Exercise 4

$$\mathbf{H}_0 : p \leq 0.25$$
$$\mathbf{H}_1 : p > 0.25$$

Hence:

$$z = \frac{105/361 - 0.25}{\sqrt{0.1875/361}} = 1.79.$$

Since this is a one-sided test, the critical value at $\alpha = 0.05$ is 1.645, and therefore we reject the null hypothesis.



# Check your knowledge:

Test your knowledge with the following multiple-choice questions.

For each question, select the correct answer. Explain your decision.

### Question 1

Which of the following is the branch of statistical inference?

a) Estimation
b) Hypothesis Testing
**c) Both a) and b)**
d) Neither a) nor b)

### Question 2

The value of an estimator is called:

a) Expectation
**b) Estimate**
c) Variance
d) None of them

### Question 3

Which of the following is not an assumption of parametric inference methods?

(a) Data is numeric.
(b) Population has a known distribution.
(c) Sample is sufficiently large.
**(d) Data is categorical.**

## Question 4

A point estimator is defined as:

a) a range of values that estimates an unknown population parameter.
b) a range of values that estimates an unknown sample statistic.
**c) a single value that estimates an unknown population parameter.**
d) a single value that estimates an unknown sample statistic.

**Explanation:** Definition: Point estimation is the use of statistics taken from one sample to estimate the value of an unknown parameter of a population.

## Question 5

Which of the following is not a characteristic for a good estimator?

a) Being unbiased.
b) Being consistent.
c) Being efficient.
**d) All of these choices are true.**

## Question 6

An unbiased estimator of a population parameter is defined as:

**a) an estimator whose expected value is equal to the parameter.**
b) an estimator whose variance is equal to one.
c) an estimator whose expected value is equal to zero.
d) an estimator whose variance goes to zero as the sample size goes to infinity.

**Explanation:** An estimator is said to be unbiased if its expected value is equal to the true value of the population.

## Question 7

An estimator is said to be consistent if:

a) it is an unbiased estimator.
b) the variance of the estimator is zero.
c) the difference between the estimator and the population parameter stays the same as the sample size grows larger.
**d) the difference between the estimator and the population parameter grows smaller as the sample size grows larger.**

## Question 8

If there are two unbiased estimators of a population parameter available, the one that has the smallest variance is said to be:

a) a biased estimator.
**b) efficient.**
c) consistent.
d) unbiased.

**Explanation:** An estimator is said to be efficient is the one that has the smallest variance is said to be relatively efficient.

## Question 9

Let $T_n$ be an estimator of θ. If E($T_n$) = θ, then

(a) $T_n$ is a sufficient estimator of θ
**(b) $T_n$ is an unbiased estimator of θ**
(c) $T_n$ is a consistent estimator of θ
(d) $T_n$ is an efficient estimator of θ

**Explanation:** An estimator is said to be unbiased if its expected value is equal to the true value of the population

## Question 10

The bias of an estimator can be:

(a) Positive

(b) Negative

(c) Zero

**(d) Any value**

**Explanation:** the bias of an estimator (or bias function) is the difference between this estimator's expected value and the true value of the parameter being estimated. A bias can assume any value. An estimator or decision rule with zero bias is called unbiased.

## Question 11

Let $X_1, X_2, \ldots X_n$ be a random sample of size *n* from a population. Then for the population variance $\sigma^2$

(a) $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an unbiased estimator.

**(b)** $\boldsymbol{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ **is a biased estimator.**

(c) $\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an unbiased estimator.

(d) None of the above.

**Explanation:** the unbiased estimator for the population variance is $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

## Question 12

The Central Limit Theorem states that the sampling distribution of the sample mean is approximately normal under certain conditions. Which of the following is a necessary condition for the Central Limit Theorem to be used?

**a) The sample size must be large.**

b) The population size must be large.

c) The population from which we are sampling must be normally distributed.

d) The population from which we are sampling must not be normally distributed.

**Explanation:** The central limit theorem says that the sampling distribution of the mean will always be normally distributed, as long as the sample size is large enough. Regardless of whether the population has a normal, Poisson, binomial, or any other distribution, the sampling distribution of the mean will be normal.

## Question 13

One year, the distribution of salaries for professional sports players had mean $1.6 million and standard deviation $0.7 million. Suppose a sample of 100 major league players was taken. Find the approximate probability that the average salary of the 100 players that year exceeded $1.1 million.

a) 0.7357
**b) approximately 1**
c) 0.2357
d) approximately 0

Explanation: calculating the probability we have:

$$
\begin{aligned}
&P(x > 1.1) = 1 - P(x < 1.1 \\
&= 1 - P(z < -12.5 \\
&= 1 - NORMSDIST(-12.5) \\
&= 1 - 0.0000 \\
&= 1.0000
\end{aligned}
$$

## Question 14

Which statement best describes a parameter?

a) A parameter is a level of confidence associated with an interval about a sample mean or proportion.
**b) A parameter is a numerical measure of a population that is almost always unknown and must be estimated.**
c) A parameter is a sample size that guarantees the error in estimation is within acceptable limits.
d) A parameter is an unbiased estimate of a statistic found by experimentation or polling.

**Explanation:** a parameter is any quantity of a statistical population that summarizes or describes an aspect of the population, such as a mean or a standard deviation. A "parameter" is to a population as a "statistic" is to a sample.

## Question 15

You are considering moving to St, Albans, and you are concerned about the average one-way commute time. Does the average one-way commute time exceed 25 minutes? You take a random sample of 50 St. Albans residents and find an average commute time of 29 minutes with a standard deviation of 7 minutes.

**a)** $\boldsymbol{H_{0:}\ \mu = 25\ \ vs\ H_{1:}\ \mu > 25}$

b) $H_{0:}\ \mu = 29\ \ vs\ H_{1:}\ \mu < 29$

c) $H_{0:}\ \mu = 25\ \ vs\ H_{1:}\ \mu \neq 25$

d) $H_{0:}\ \mu = 29\ \ vs\ H_{1:}\ \mu > 29$

e) $H_{0:}\ \mu = 25\ \ vs\ H_{1:}\ \mu < 25$

**Explanation:** correct answer is a) because you are interested in knowing of the average one-way commute time exceed 25 minutes (related to the population). You know that in your sample the average commute time is 29.

## Question 16

Suzie has installed a new spam blocker program on her email. She used to receive an average of 20 spam emails a day. Is the new program working?

**a)** $\boldsymbol{H_{0:}\ \mu = 20\ \ vs\ H_{1:}\ \mu < 20}$

b) $H_{0:}\ \mu > 20\ \ vs\ H_{1:}\ \mu = 20$

c) $H_{0:}\ \mu = 20\ \ vs\ H_{1:}\ \mu > 20$

d) $H_{0:}\ \mu = 20\ \ vs\ H_{1:}\ \mu \neq 20$

e) Not enough information is given

**Explanation:** correct answer is a). if the new program is working, you will expect to receive less than 20 spam emails.

## Question 17

We reject the null hypothesis if the sample data falls in the

a) critical region
**b) rejection region**
c) acceptance region
d) None of these

## Question 18

The statistic based on whose value the null hypothesis is rejected or notis called

**a) the test statistic**
b) the critical value
c) both (a) and (b)
d) Neither (a) nor (b).

**Explanation:** A test statistic is a statistic (a quantity derived from the sample) used in statistical hypothesis testing. A hypothesis test is typically specified in terms of a test statistic. It is selected or defined in such a way as to quantify, within observed data, behaviours that would distinguish the null from the alternative hypothesis,

## Question 19

The value of the test statistic which separates the rejection region and acceptance region is called:

a) test statistic value
b) level of significance
**c) critical value**
d) None of the above

**Explanation:** the critical values of a statistical test are the boundaries of the acceptance region of the test. The acceptance region is the set of values of the test statistic for which the null hypothesis is not rejected.

**Question 20**

In the test for proportion $H_{0:}\, p_1 = p_2 \;\; vs \; H_{1:}\, p_1 \neq p_2,$ the best critical region is given by

a) $Z > z_\alpha$
b) $Z > -z_\alpha$
c) $Z < z_\alpha$
**d) None of the above**

**Explanation:** this is a two sided test. Correct answer is $Z < -z_{\alpha/2}\; and\; Z > z_{\alpha/2}$

**Question 21**

A large university is interested in learning about the average time it takes students to travel to campus. The university sampled 238 students and asked each to provide the amount of time they spent traveling to campus. This variable, travel time, was then used conduct a test of hypothesis. The goal was to determine if the average travel time of all the university's students differed from 20 minutes. You know that the average travel time in your sample is 23.243 and the population standard deviation is 1.3133. Conducting a test at 1% significant level.

What conclusion can be made? When testing at $\alpha$ = 0.01…

a) …there is sufficient evidence to indicate that the average travel time of all students exceeds 20 minutes.
**b) …there is sufficient evidence to indicate that the average travel time of all students is equal to 20 minutes.**
c) …there is insufficient evidence to indicate that the average travel time of all students is equal to 20 minutes.
d) …there is insufficient evidence to indicate that the average travel time of all students exceeds 20 minutes.

**Explanation:** $H_{0:}\, \mu = 20 \;\; vs \; H_{1:}\, \mu \neq 20$

$$z = \frac{23.234 - 20}{1.3133} = 2.4693$$

The critical values obtain from the tables are -2.58, +2.58. They are calculated as P(Z<0.005)

Since 2.4693 < 2.58, then we fail to reject H0. Correct answer is b)

## Question 22

Repeat the exercise in question 17. Now conduct the test at 5% significant level.

What conclusion can be made? When testing at $\alpha$ = 0.05…

a) **…there is sufficient evidence to indicate that the average travel time of all students exceeds 20 minutes.**
b) …there is sufficient evidence to indicate that the average travel time of all students is equal to 20 minutes.
c) …there is insufficient evidence to indicate that the average travel time of all students is equal to 20 minutes.
d) …there is insufficient evidence to indicate that the average travel time of all students exceeds 20 minutes.

**Explanation:** $H_{0:}\ \mu = 20\ \ vs\ H_{1:}\ \mu \neq 20$

$$z = \frac{23.234 - 20}{1.3133} = 2.4693$$

The critical values obtain from the tables are -1.96, +1.96. They are calculated as P(Z<0.025)

Since 2.4693 > 1.96, then we reject H0. Correct answer is a)

## Question 23

A national organization has been working with utilities throughout the nation to find sites for large wind machines that generate electricity. Wind speeds must average more than 19 miles per hour (mph) for a site to be acceptable. Recently, the organization conducted wind speed tests at a particular site. Based on a sample of n = 45 wind speed recordings (taken at random intervals), the wind speed at the site averaged $\bar{x}$ = 19.9 mph, with a standard deviation of 4.5 mph. To determine whether the site meets the organization's requirements, consider the test, $H_{0:}\ \mu = 19\ \ vs\ H_{1:}\ \mu > 19$, where μ is the true mean wind speed at the site and $\alpha$ = 0.10. Suppose the value of the test statistic were computed to be 1.34. State the conclusion.

a) At $\alpha$ = 0.10, there is evidence to conclude the true mean wind speed at the site does not exceeds 19 mph.
b) **At $\alpha$ = 0.10, there is evidence to conclude the true mean wind speed at the site exceeds 19 mph.**

c) We don't have enough information to reach a conclusion.

Explanation: The critical value obtained from the standard normal tables is 1.285. This exercise is based on a one-sided test. Therefore, we have one rejection area (on the right tail). The reason why the rejection area is on the right tail is because the alternative hypothesis is $H_1: \mu > 19$. Now we need to find the critical value ($z_\alpha$). The critical value can be found using statistical tables.

Critical value: $P(Z > z_\alpha) = 0.10$. Otherwise, you can look for the probability $P(Z < z_\alpha) = 0.90$.

Rejection rule: reject the null hypothesis if the test statistic is greater than the critical value. Since we have z-stat(1.34) > critical value(1.285), we can reject the null hypothesis and conclude that the true mean wind speed at the site exceeds 19 mph.

**Question 24**

Statistical quality control is based on the theory of

a) probability
b) sampling
**c) both (a) and (b)**
d) neither (a) nor (b)

**Explanation:** Statistical quality control (SQC) is the application of statistical methods to monitor and control the quality of a production process. SQC is defined as the technique of applying statistical methods based on the theory of probability and sampling to establish quality standard and to maintain it in the most economical manner.

**Question 25**

The manager of a grocery store has taken a random sample of 100 customers. The average length of time it took the customers in the sample to check out was 2.9 minutes with a population standard deviation of 0.07 minutes. Do you reject the claim that mean waiting time of all customers is significantly less than 3 minutes. Use α = 0.05.

a) Yes. The mean waiting time of all customers is more than 3 minutes.
**b) No. The mean waiting time of all customers is less than 3 minutes.**

**Explanation:** $H_{0:}\ \mu \leq 3\ \ vs\ H_{1:}\ \mu > 3$

$$z = \frac{2.9 - 3}{0.07} = 1.4286$$

The critical values obtain from the tables are -1.96, +1.96. They are calculated as P(Z<0.025)

Since 1.4286 < 1.96, then we fail to reject H0. Correct answer is b)