One-Sample Procedure

The Federal Aviation Administration (FAA) requires that environmental documents address noise impact around airports using an impact threshold of Day Night Average Sound Level (DNL) 65(dBa). A **sample of 112 residential homes** in the vicinity of a major airport were selected and the average DNL noise level of these residents were measured with the owner's permission over a 1-month period. The data is used to assess if there is evidence at an **0.06 significance level** that the DNL **exceeds 65(dBa)**. The **sample mean and sample standard deviation** for these **112 observations** for the average DNL noise level was found to be **66.00179(dBa) and 6.717169(dBa) respectively**.

a) Select the letter corresponding to the correct choice for the first step of the four-step hypothesis test process.

[A] The parameter of interest is the sample Day Night Average Sound Level (DNL) **"** $\boldsymbol{\bar{x}_{\mathrm{DNL}}}$**"**of the 112 residential homes sampled from around this major airport; in units of dBA.

[B] The parameter of interest is the sample Day Night Average Sound Level (DNL) **"** $\boldsymbol{\bar{x}_{\mathrm{DNL}}}$**"** of the neighborhood of residential homes around this major airport; in units of dBA.

[C] The parameter of interest is the true Day Night Average Sound Level (DNL) **"** $\boldsymbol{\bar{x}_{\mathrm{DNL}}}$**"** of the neighborhood of residential homes around this major airport; in units of dBA.

[D] The parameter of interest is the sample Day Night Average Sound Level (DNL) "$\boldsymbol{\mu}_{\mathbf{DNL}}$" of the 112 residential homes sampled from around this major airport; in units of dBA.

[E] The parameter of interest is the true Day Night Average Sound Level (DNL) "$\boldsymbol{\mu}_{\mathbf{DNL}}$" of the neighborhood of residential homes around this major airport; in units of dBA.

[F] The parameter of interest is the sample Day Night Average Sound Level (DNL) "$\boldsymbol{\mu}_{\mathbf{DNL}}$" of the neighborhood of residential homes around this major airport; in units of dBA.

b) Select the letter corresponding to the most appropriate choice for the second step **(State the Hypothesis)** of the four-step hypothesis test process.

[A] $\mathbf{H_0}$: $\boldsymbol{\mu}_{\mathbf{DNL}} \leq \mathbf{65}$ $(\mathbf{dBa})$, $\mathbf{H}_{\boldsymbol{a}}$: $\boldsymbol{\mu}_{\mathbf{DNL}} > \mathbf{65}$ $(\mathbf{dBa})$

[B] $\mathbf{H_0}$: $\boldsymbol{\mu}_{\mathbf{DNL}} \geq \mathbf{65}$ $(\mathbf{dBa})$, $\mathbf{H}_{\boldsymbol{a}}$: $\boldsymbol{\mu}_{\mathbf{DNL}} < \mathbf{65}$ $(\mathbf{dBa})$

[C] $\mathbf{H_0}$: $\boldsymbol{\mu}_{\mathbf{DNL}} = \mathbf{65}$ $(\mathbf{dBa})$, $\mathbf{H}_{\boldsymbol{a}}$: $\boldsymbol{\mu}_{\mathbf{DNL}} \neq \mathbf{65}$ $(\mathbf{dBa})$

[D] $\mathbf{H_0}$: $\boldsymbol{\mu}_{\mathbf{DNL}} \neq \mathbf{65}$ $(\mathbf{dBa})$, $\mathbf{H}_{\boldsymbol{a}}$: $\boldsymbol{\mu}_{\mathbf{DNL}} = \mathbf{65}$ $(\mathbf{dBa})$

[E] $\mathbf{H_0}$: $\boldsymbol{\mu}_{\mathbf{DNL}} = \mathbf{65}$ $(\mathbf{dBa})$, $\mathbf{H}_{\boldsymbol{a}}$: $\boldsymbol{\mu}_{\mathbf{DNL}} < \mathbf{65}$ $(\mathbf{dBa})$

[F] $\mathbf{H_0}$: $\boldsymbol{\mu}_{\mathbf{DNL}} = \mathbf{65}$ $(\mathbf{dBa})$, $\mathbf{H}_{\boldsymbol{a}}$: $\boldsymbol{\mu}_{\mathbf{DNL}} > \mathbf{65}$ $(\mathbf{dBa})$

c) Which of the following would provide a result consistent with the hypothesis you selected above:

[A] **Confidence Interval**

[B] Lower Confidence Bound

[C] Upper Confidence Bound

d) What is the appropriate **confidence level** for the **interval or bound** that you selected above?

Please write your answer using **4 decimal places** and write it as **decimal value not** as a **percentage**.

For example, if it was a **95.55% confidence level**, you would enter **0.9555**. You may drop any trailing zeros if you do not need the full 4 decimal places.

$$\boldsymbol{C = 100 \times (1 - \alpha) = 94\%}$$

$\mathbf{0.94}$ **<-answer**

e) An FAA researcher has determined from prior studies that the average DNL noise level in neighborhoods around airports is **known** to have a **standard deviation of 4(dBa)**.

Select the **correct critical value** from the **table below** for the **confidence interval** or **bound** that is **consistent** with your **stated hypothesis** and **confidence level**.

Note that more information is provided in the table than is needed to answer the question.

> qt(0.06, df = 111, lower.tail = FALSE) [1] 1.566834	> qt(0.06/2, df = 111, lower.tail = FALSE) [1] 1.90021
> qnorm(0.06, lower.tail = FALSE) [1] 1.554774	> qnorm(0.06/2, lower.tail = FALSE) [1] 1.880794

Your answer should be given using the full value of the critical value. Do not round and make sure you use the full value in subsequent questions.

$\sigma = 4$ is known so we use normal distribution.

> qnorm(0.06, lower.tail = FALSE)

[1] 1.554774

f) Use the **critical value** you just obtained and the **population standard deviation** of **4 dBa** to **calculate** the **confidence interval** or **bound** that is consistent with your stated hypothesis.

Your answer must be given in **interval notation** and any **numbers** should be **correctly rounded** to **four decimal places**.

We will use confidence lower bound since the alternative is greater than.

$$\left(\bar{x}_{\text{DNL}} - z_{0.06}\frac{\sigma_{\text{DNL}}}{\sqrt{n}}, \infty\right)$$

$$\left(66.00179 - 1.554774\frac{4}{\sqrt{112}}, \infty\right)$$

$$(65.4141, \infty)$$

g) Using the confidence interval or bound that you computed above determine the conclusion of the hypothesis test.

Would you reject the null hypothesis? Select the most appropriate option from the choices below.

<u>**Confidence Upper Bound Choices**</u>

[A] The **confidence upper bound** value is **greater than** the **null value of 65 (dBa)**. Therefore, we **would reject** the **null**

hypothesis that the **true DNL** is **at least 65 (dBa)** and conclude that there **is evidence** at the **0.06 significance level** that the **true DNL** is **less than 65 (dBa)**.

[B] The **confidence upper bound** value is **greater than** the **null value of 65 (dBa)**. Therefore, we **would reject** the **null hypothesis** that the **true DNL** is **at least 65 (dBa)** and conclude that there is **no evidence** at the **0.06 significance level** that the **true DNL** is **less than 65 (dBa)**.

[C] The **confidence upper bound** value is **greater than** the **null value of 65 (dBa)**. Therefore, we **would not reject** the **null hypothesis** that the **true DNL** is **at least 65 (dBa)** and conclude that there **is evidence** at the **0.06 significance level** that the **true DNL** is **less than 65 (dBa)**.

[D] The **confidence upper bound** value is **greater than** the **null value of 65 (dBa)**. Therefore, we **would not reject** the **null hypothesis** that the **true DNL** is **at least 65 (dBa)** and conclude that there is **no evidence** at the **0.06 significance level** that the **true DNL** is **less than 65 (dBa)**.

<u>Confidence Interval Choices</u>

[E] The **null value** of **65 (dBa)** is **not within** the **confidence interval**. Therefore, we would **reject** the **null hypothesis** that the **true DNL** is **65 (dBa)** and conclude that there **is evidence** at the **0.06 significance level** that the **true DNL** is **not equal** to **65 (dBa)**.

[F] The **null value** of **65 (dBa)** is **not within** the **confidence interval**. Therefore, we would **not reject** the **null hypothesis** that the **true DNL** is **65 (dBa)** and conclude that there **is no evidence** at the **0.06 significance level** that the **true DNL** is **not equal** to **65 (dBa)**.

[G] The **null value** of **65 (dBa)** is **within** the **confidence interval**. Therefore, we would **not reject** the **null hypothesis**

that the **true DNL** is **65 (dBa)** and conclude that there **is evidence** at the **0.06 significance level** that the **true DNL** is **not equal** to **65 (dBa)**.

[H] The **null value** of **65 (dBa)** is **within** the **confidence interval**. Therefore, we would **reject** the **null hypothesis** that the **true DNL** is **65 (dBa)** and conclude that there **is no evidence** at the **0.06 significance level** that the **true DNL** is **not equal** to **65 (dBa)**.

Confidence Lower Bound Choices

[I] The **confidence lower bound** value is **greater than** the **null value of 65 (dBa)**. Therefore, we would **not reject** the **null hypothesis** that the **true DNL** is **at most 65 (dBa)** and conclude that there **is evidence** at the **0.06 significance level** that the **true DNL** is **greater than 65 (dBa)**.

[J] The **confidence lower bound** value is **greater than** the **null value of 65 (dBa)**. Therefore, we would **not reject** the **null hypothesis** that the **true DNL** is **at most 65 (dBa)** and conclude that there is **no evidence** at the **0.06 significance level** that the **true DNL** is **greater than 65 (dBa**

[K] The **confidence lower bound** value is **greater than** the **null value of 65 (dBa)**. Therefore, we would **reject** the **null hypothesis** that the **true DNL** is **at most 65 (dBa)** and conclude that there is **evidence** at the **0.06 significance level** that the **true DNL** is **greater than 65 (dBa)**.

[L] The **confidence lower bound** value is **greater than** the **null value of 65 (dBa)**. Therefore, we would **reject** the **null hypothesis** that the **true DNL** is **at most 65 (dBa)** and conclude that there is **no evidence** at the **0.06 significance level** that the **true DNL** is **greater than 65 (dBa)**.

h) Another researcher has identified that the design of the neighborhoods around this airport differs from those around typical airports and it would **not be appropriate** to **assume** that the **standard deviation is 4(dBa)**. Would the **critical value** of the **confidence interval** or **bound change**?

Select the letter of the most appropriate answer from the choices below.

> **[A] Yes**, the **critical value would change**; in fact, it would be **larger** than what we found previously. The need for a new critical value is **due to the characteristics of the neighborhood** forcing the **selection** of a **new larger critical value**.
>
> **[B] Yes**, the **critical value would change**; in fact, it would be **smaller** than what we found previously. The need for a new critical value is **due to the characteristics of the neighborhood** forcing the selection of a **new smaller critical value**.
>
> **[C] Yes**, the **critical value would change**; in fact, it would be **larger** than what we found previously. The need for a **new critical value** is **due** to the **added uncertainty** of having to **estimate** the **standard deviation** from the data.
>
> **[D] Yes**, the **critical value would change**; in fact, it would be **smaller** than what we found previously. The need for a **new critical value** is **due** to the **added uncertainty** of having to **estimate** the **standard deviation** from the data.
>
> **[E] No**, the **critical value would not change**. The choice of the critical value is **not dependent** on the **knowledge** of the **population standard deviation**.
>
> **[F] No**, the **critical value would not change**. The **only thing** that **would change** is the **standard deviation** as we now need to **estimate** it from the data.
>
> **[G] No**, the **critical value would not change**. The **critical value** is solely **determined** by the **sample size** and **significance**

level, not by whether the **standard deviation is known** or **estimated**.

[H] No, the **critical value would not change**. The **critical value** is a **constant** for a **given significance level** and **degrees of freedom**, not by whether the **standard deviation** is **known** or **estimated**.

i) What practical answer would you give to the FAA about the noise level around this particular major airport? Differences greater than 1.5(dBa) higher than the **impact threshold DNL** of **65(dBa)** are **considered problematic** and **require significant investment** to reduce the noise level. Use the **confidence interval or bound** found previously to provide a **practicality assessment** for the FAA. You may continue to assume the **population standard deviation** is **known**, regardless of your conclusion in the previous part.

You must report your **conclusion, difference,** and **effect size**.

Write your answer as follows: conclusion (yes or no), difference, effect size. Round your number to 2 decimal places for simplicity.

For **example**, if you concluded **yes** there is **practical significance (investment is required to reduce the noise level)**, and found a difference of **3.2 dBa**, with an **effect size** of **0.81** you would write your answer as: --> **yes,3.2,0.81**

Confidence lower bound value → $\mathbf{65.4141}$

Null value → 65

Difference → $\mathbf{65.4141 - 65 \approx 0.41}$

Effect Size → $\mathbf{\frac{0.41}{4} \approx 0.1}$

Answer → no,0.41,0.1

Two-Sample Procedure

A university department is conducting research to enhance the undergraduate learning environment. The department aims to investigate whether there is a **positive** impact on learning outcomes when the emphasis of a course is shifted from **traditional lecture-based instruction (T)** to new **AI-enhanced instruction (A)**.

To accomplish this, the department separated its **56 undergraduate courses randomly** into two groups: **28 courses** would be taught using the new **AI-enhanced instruction**, and the **remaining 28 courses** would be taught using the **traditional lecture-based** instruction. At the end of the semester, they collected the **average final grades from each course**.

The summary statistics of the semester are provided in the table below.

Course Type	n	$\bar{x}$	s
A	28	78.82809	2.810766
T	28	77.59551	3.182277
A-T	28	1.232584	4.161893

a) Which analysis method is appropriate for the given situation: **two-sample independent** or **two-sample paired**?

Select the correct approach and the most appropriate reasoning from the choices below.

Two-Sample Independent Choices

[A] The situation calls for a **two-sample independent analysis**. The department randomly assigned 28 courses to receive the AI-enhanced instruction and the remaining 28 courses to receive traditional lecture-based instruction. Since the courses were assigned independently and there is no common factor to pair on, a two-sample independent analysis is appropriate.

[B] The situation calls for a **two-sample independent analysis** since the department assigned students to different courses at random, ensuring that each course had a unique group of students. This random assignment of students to courses means there are no shared characteristics between the groups that could be used for pairing, thus eliminating any potential pairing factors.

[C] The situation calls for a **two-sample independent analysis**. The average grades were compared across the two different teaching methods without considering any pairing between courses. This approach treats each group of courses as separate and independent, assuming there are no shared characteristics that would link courses together for a paired analysis.

Two-Sample Paired Choices

[D] The situation calls for a **two-sample paired analysis** because each AI-enhanced course is paired with a lecture-based course based on similar content and difficulty. This matching ensures that each pair of courses is directly comparable in terms of subject matter and instructional level.

[E] The situation calls for a **two-sample paired analysis**. The courses were matched based on student demographics and prior performance. By pairing courses with similar student populations, the analysis accounts for differences in student backgrounds and abilities.

[F] The situation calls for a **two-sample paired** analysis because the same students took both types of courses, providing a direct comparison. This approach allows for a within-subject comparison, where each student serves as their own control, reducing variability.

b) The department will conduct a statistical hypothesis for this investigation at a **0.02 significance level**. Select the most appropriate hypothesis for the given statistical question from the choices below.

Let **A** denote the enhanced AI-based instruction, and **T** denote the traditional lecture-based instructional method and $\boldsymbol{\mu}$ denote the true average grades.

[A] $\mathbf{H_0}: \boldsymbol{\mu_A - \mu_T \geq 0}$, $\mathbf{H_a}: \boldsymbol{\mu_A - \mu_T < 0}$

[B] $\mathbf{H_0}: \boldsymbol{\mu_A - \mu_T \leq 0}$, $\mathbf{H_a}: \boldsymbol{\mu_A - \mu_T > 0}$

[C] $\mathbf{H_0}: \boldsymbol{\mu_A - \mu_T = 0}$, $\mathbf{H_a}: \boldsymbol{\mu_A - \mu_T \neq 0}$

Let the difference **D** be defined as **D = A - T**.

[D] $\mathbf{H_0}$: $\boldsymbol{\mu}_\mathbf{D} \geq \mathbf{0}$, $\mathbf{H}_{\boldsymbol{a}}$: $\boldsymbol{\mu}_\mathbf{D} < \mathbf{0}$

[E] $\mathbf{H_0}$: $\boldsymbol{\mu}_\mathbf{D} \leq \mathbf{0}$, $\mathbf{H}_{\boldsymbol{a}}$: $\boldsymbol{\mu}_\mathbf{D} > \mathbf{0}$

[F] $\mathbf{H_0}$: $\boldsymbol{\mu}_\mathbf{D} = \mathbf{0}$, $\mathbf{H}_{\boldsymbol{a}}$: $\boldsymbol{\mu}_\mathbf{D} \neq \mathbf{0}$

c) Calculate the appropriate test statistic. **(Use 4 decimal places in your answer)**

$$t_{\mathrm{TS}} = \frac{(\bar{x}_{\mathrm{A}} - \bar{x}_{\mathrm{T}})}{\sqrt{\frac{s_{\mathrm{A}}^2}{n_{\mathrm{A}}} + \frac{s_{\mathrm{T}}^2}{n_{\mathrm{T}}}}} = \frac{78.82809 - 77.59551}{\sqrt{\frac{2.810766^2}{28} + \frac{3.182277^2}{28}}} = 1.536133$$

d) Indicate which of the options below is **most likely** to be the correct R code for directly computing the appropriate p-value for the hypothesis test?

Let **TS** denote the assumed correctly computed test statistic value you computed above.

[A] pnorm(**TS**, lower.tail= TRUE)

[B] 2*pnorm(|**TS**|, lower.tail=FALSE)

[C] pnorm(**TS**, lower.tail=FALSE)

[D] pt(**TS**, df= 27 , lower.tail=TRUE)

[E] 2 * pt(|**TS**|, df= 27 lower.tail=FALSE)

[F] pt(**TS**, df= 27, lower.tail=FALSE)

[G] pt(**TS**, df= 53.18867, lower.tail=TRUE)

[H] 2 * pt(|**TS**|, df= 53.18867, lower.tail=FALSE)

[I] pt(**TS**, df= 53.18867, lower.tail=FALSE)

e) Assume that the following interval and bounds are computed correctly using the data and at a confidence level of 0.98.

Choose the appropriate confidence bound/interval for this situation.

[A] 98 percent confidence interval: **-0.4567313 Inf**

[B] 98 percent confidence interval: **-Inf 2.921891**

[C] 98 percent confidence interval: **-0.691976 3.157136**

f) Select an appropriate interpretation of your confidence interval or bound.

[A] I am **98% confident** that the **difference** in the **true means** between **AI-enhanced instruction** and **traditional lecture-based instruction** is **less than -0.4567313**.

[B] I am **98% confident** that the **difference** in the **true means** between **AI-enhanced instruction** and **traditional lecture-based instruction** is **greater than -0.4567313**.

[C] I am **98% confident** that the **difference** in the **true means** between **AI-enhanced instruction** and **traditional lecture-based instruction** is **less than 2.921891**.

[D] I am **98% confident** that the **difference** in the **true means** between **AI-enhanced instruction** and **traditional lecture-based instruction** is **greater than 2.921891**.

[E] I am **98% confident** that the **difference** in the **true means** between **AI-enhanced instruction** and **traditional lecture-based instruction** is **captured** by the interval **(-0.691976, 3.157136)**.

[F] There is a **98% probability** that the **difference** in the **true means** between **AI-enhanced instruction** and **traditional lecture-based instruction** is **within** the interval **(-0.691976, 3.157136)**.

g) Based **solely** on the **appropriate confidence interval/bound** that you selected above, what is the most **appropriate conclusion** for the **hypothesis test** from the choices below.

[A] Do not reject the null hypothesis. The data **does not provide** sufficient **evidence** to **reject** the **null hypothesis** at the **significance level (**α=0.02**)**, indicating **no significant difference** between the instructional methods.

[B] Reject the null hypothesis. The data **does provide** sufficient **evidence** to **reject** the **null hypothesis** at the **significance level (**α=0.02**)**, indicating a **significant difference** between the instructional methods.

[C] Do not reject the null hypothesis. The **p-value** was found to be **greater** than the **significance level (**α=0.02**)**, indicating that we **do not have sufficient evidence** to **reject** the **null hypothesis** and suggest a **significant difference** between the instructional methods.

[D] Reject the null hypothesis. The **p-value** was found to be **less** than the **significance level (**α=0.02**)**, indicating that we **do have sufficient evidence** to **reject** the **null hypothesis** and suggest a **significant difference** between the instructional methods.

[E] Do not reject the null hypothesis. Since the constructed **98% confidence interval** or **bound** is complementary to the hypothesis test **(with** α+C=1**),** is in the **direction** of the **alternative hypothesis**, and **contains 0**, we would **not reject** the **null hypothesis** and therefore we **do not have sufficient evidence** to suggest that the AI-enhanced instruction has a positive impact on learning outcomes in comparison to traditional based instruction.

[F] Reject the null hypothesis. Since the constructed **98% confidence interval** or **bound** is complementary to the hypothesis test **(with** α+C=1**),** is in the **direction** of the **alternative hypothesis**, and **contains 0**, we would **reject** the **null hypothesis** and therefore we **have sufficient evidence** to suggest the AI-enhanced instruction has a positive impact on learning outcomes in comparison to traditional based instruction.

Central Limit Theorem

Suppose the **heights** of **adult oak trees** in a large forest follow a **slightly positively skewed distribution** with an **average height** of **60 feet** and a **standard deviation** of **8 feet**. A researcher randomly selects a **sample** of **36 adult oak trees** from this **population**.

a) What is the **distribution** of the **mean heights** for **random samples** of **36 adult oak trees**? Clearly specify the **name** of the **distribution** and its **parameter(s)**.

Let $\mathbf{T}$ denote the height of adult oak trees in this large forest and $\overline{\mathbf{T}}$ denote the random variable measuring the average height of a random sample of 36 adult oak trees in this large forest.

$$\overline{\mathrm{T}} \sim N\left(\mu_{\overline{\mathrm{T}}} = 60, \sigma_{\overline{\mathrm{T}}} = \frac{8}{\sqrt{36}} = \frac{4}{3} = 1.3333\right)$$

b) Given that the **mean height** of the **36 adult oak trees** is known to be **at least 59 feet**, what is the **probability** that the **mean height** is **more than 62 feet**?

Clearly set up the probability to be calculated and show the mathematical steps required to obtain the probability.

You may use the following R output in your calculations.

pnorm(-0.75, lower.tail = TRUE) 0.2266274	pnorm(-0.75, lower.tail = FALSE) 0.7733726
pnorm(-0.125, lower.tail = FALSE) 0.4502618	pnorm(-0.125, lower.tail = FALSE) 0.5497382
pnorm(0.25, lower.tail = TRUE) 0.5987063	pnorm(0.25, lower.tail = FALSE) 0.4012937
pnorm(1.5, lower.tail = TRUE) 0.9331928	pnorm(1.5, lower.tail = FALSE) 0.0668072

$$P(\overline{\mathrm{T}} > 62 | \overline{\mathrm{T}} \geq 59) = \frac{P(\overline{\mathrm{T}} > 62)}{P(\overline{\mathrm{T}} \geq 59)} = \frac{P\left(Z > \frac{62-60}{4/3}\right)}{P\left(Z > \frac{59-60}{4/3}\right)}$$

$$= \frac{P(Z > 1.5)}{P(Z > -0.75)} = \frac{0.0668072}{0.7733726} = 0.0864$$

c) What is the **mean height value** that **separates** the **top 1%** of the mean height of a sample of 36 adult oak trees? Show your work.

You may use the following R output in your calculations.

qnorm(0.01/2, lower.tail = FALSE) 2.575829	qnorm(0.01, lower.tail = FALSE) 2.326348
qnorm(0.05/2, lower.tail = FALSE) 1.959964	qnorm(0.05, lower.tail = FALSE) 1.644854

Top 1% standard normal quantile →

qnorm(0.01, lower.tail = FALSE)

$$z^* = 2.326348$$

Move to the sampling distribution for the sample mean of adult oak tree heights for samples of size 36.

$$\bar{t} = \mu_{\bar{T}} + z^* \sigma_{\bar{T}} = 60 + 2.326348 \times \frac{4}{3} = 63.1018$$

Central Limit Theorem Version 2

Suppose the **diameters** of **adult pine trees** in a large forest follow a **slightly positively skewed distribution** with an **average diameter** of **30 inches** and a **standard deviation** of **5 inches**. A researcher randomly selects a sample of **64 adult pine trees** from this **population**.

a) What is the **distribution** of the **average diameter** of a **random sample** of **64 adult pine trees**? Clearly specify the **name** of the **distribution** and its **parameter(s)**.

Let $\mathbf{T}$ denote the diameter of adult pine trees in this large forest and $\overline{\mathbf{T}}$ denote the random variable measuring the average diameter of a random sample of 64 adult pine trees in this large forest.

$$\overline{\mathrm{T}} \sim N\left(\mu_{\overline{\mathrm{T}}} = 30, \sigma_{\overline{\mathrm{T}}} = \frac{5}{\sqrt{64}} = \frac{5}{8} = 0.625\right)$$

b) Given that the **mean diameter** of the **64 adult pine trees** is known to be **at most 31 inches**, what is the probability that the **mean diameter** is **less than 28 inches**?

Clearly set up the probability to be calculated and show the mathematical steps required to obtain the probability.

You may use the following R output in your calculations.

pnorm(-3.2, lower.tail = TRUE) 0.0006871379	pnorm(-3.2, lower.tail = FALSE) 0.9993129
pnorm(-0.4, lower.tail = FALSE) 0.3445783	pnorm(-0.4, lower.tail = FALSE) 0.6554217
pnorm(0.2, lower.tail = TRUE) 0.5792597	pnorm(0.2, lower.tail = FALSE) 0.4207403
pnorm(1.6, lower.tail = TRUE) 0.9452007	pnorm(1.6, lower.tail = FALSE) 0.05479929

$$P(\overline{\mathrm{T}} < 28 | \overline{\mathrm{T}} \leq 31) = \frac{P(\overline{\mathrm{T}} < 28)}{P(\overline{\mathrm{T}} \leq 31)} = \frac{P\left(Z < \frac{28-30}{5/8}\right)}{P\left(Z < \frac{31-30}{5/8}\right)}$$

$$= \frac{P(Z < -3.2)}{P(Z < 1.6)} = \frac{0.0006871379}{0.9452007} = 0.0007$$

c) What is the **mean diameter value** that separates the **top 5%** of the **mean diameters** of a sample of **64 adult pine trees**? Show your work.

You may use the following R output in your calculations.

qnorm(0.01/2, lower.tail = FALSE) 2.575829	qnorm(0.01, lower.tail = FALSE) 2.326348
qnorm(0.05/2, lower.tail = FALSE) 1.959964	qnorm(0.05, lower.tail = FALSE) 1.644854

Top 5% standard normal quantile →

qnorm(0.05, lower.tail = FALSE)

$$z^* = 1.644854$$

Move to the sampling distribution for the sample mean of adult pine tree diameters for samples of size 64.

$$\bar{t} = \mu_{\bar{T}} + z^* \sigma_{\bar{T}} = 30 + 1.644854 \times \frac{5}{8} = 31.028$$

Central Limit Theorem Version 3

Suppose the **wingspan** of **adult eagles** in a national park follow a **slightly positively skewed distribution** with an **average wingspan** of **7 feet** and a **standard deviation** of **3.5 feet**. A researcher randomly selects a **sample** of **49 adult eagles** from this population.

a) What is the **distribution** of the **mean wingspans** of a **random sample** of **49 adult eagles**? Clearly specify the **name** of the **distribution** and its **parameter(s)**.

Let $\mathbf{W}$ denote the wingspan of adult eagles in this national park and $\bar{\mathbf{W}}$ denote the random variable measuring the average wingspan of a random sample of 49 adult pine trees in this large forest.

$$\bar{\mathbf{W}} \sim N\left(\mu_{\bar{W}} = 7, \sigma_{\bar{W}} = \frac{3.5}{\sqrt{49}} = \frac{3.5}{7} = 0.5\right)$$

d) Given that the **mean wingspan** of the **49 adult eagles** is known to be **at least 6.625 feet**, what is the probability that the **mean wingspan** is **more than 7.75 feet**?

Clearly set up the probability to be calculated and show the mathematical steps required to obtain the probability.

You may use the following R output in your calculations.

pnorm(-0.75, lower.tail = TRUE) 0.2266274	pnorm(-0.75, lower.tail = FALSE) 0.7733726
pnorm(-0.11, lower.tail = FALSE) 0.4562047	pnorm(-0.11, lower.tail = FALSE) 0.5437953
pnorm(0.21, lower.tail = TRUE) 0.5831662	pnorm(0.21, lower.tail = FALSE) 0.4168338
pnorm(1.5, lower.tail = TRUE) 0.9331928	pnorm(1.5, lower.tail = FALSE) 0.0668072

$$P(\bar{W} > 7.75 | \bar{W} \geq 6.625) = \frac{P(\bar{W} > 7.75)}{P(\bar{W} \geq 6.625)} = \frac{P\left(Z < \frac{7.75 - 7}{1/2}\right)}{P\left(Z < \frac{6.625 - 7}{1/2}\right)}$$

$$= \frac{P(Z > 1.5)}{P(Z > -0.75)} = \frac{0.0668072}{0.7733726} = 0.0864$$

b) What is the **mean wingspan value** that separates the **top 5%** of the **mean wingspans** of a **sample** of **49 adult eagles**? Show your work.

You may use the following R output in your calculations.

qnorm(0.01/2, lower.tail = FALSE) 2.575829	qnorm(0.01, lower.tail = FALSE) 2.326348
qnorm(0.05/2, lower.tail = FALSE) 1.959964	qnorm(0.05, lower.tail = FALSE) 1.644854

Top 5% standard normal quantile →

qnorm(0.05, lower.tail = FALSE)

$$z^* = 1.644854$$

Move to the sampling distribution for the sample mean of adult wingspans for samples of size 49.

$$\bar{w} = \mu_{\bar{W}} + z^* \sigma_{\bar{W}} = 7 + 1.644854 \times \frac{1}{2} = 7.8225$$

Central Limit Theorem Version 4

Suppose the **lengths** of **adult salmon** in a river follow a **slightly positively skewed distribution** with an **average length** of **28 inches** and a **standard deviation of 5 inches**. A researcher randomly selects a **sample** of **64 adult salmon** from this population.

a) What is the distribution of the mean lengths of a **random sample** of **64 adult salmon**? Clearly specify the **name** of the **distribution** and its **parameter(s)**.

Let $\boldsymbol{S}$ denote the length of adult salmon in this river and $\bar{\mathbf{S}}$ denote the random variable measuring the average length of a random sample of 64 adult salmon in this large river.

$$\bar{\mathbf{S}} \sim N\left(\mu_{\bar{\mathbf{S}}} = 28, \sigma_{\bar{\mathbf{S}}} = \frac{5}{\sqrt{64}} = \frac{5}{8} = 0.625\right)$$

b) Given that the **mean length** of the **64 adult salmon** is known to be at **most 29 inches**, what is the probability that the **mean length** is **less than 26 inches**?

Clearly set up the probability to be calculated and show the mathematical steps required to obtain the probability.

You may use the following R output in your calculations.

pnorm(-3.2, lower.tail = TRUE) 0.0006871379	pnorm(-3.2, lower.tail = FALSE) 0.9993129
pnorm(-0.4, lower.tail = FALSE) 0.3445783	pnorm(-0.4, lower.tail = FALSE) 0.6554217
pnorm(0.2, lower.tail = TRUE) 0.5792597	pnorm(0.2, lower.tail = FALSE) 0.4207403
pnorm(1.6, lower.tail = TRUE) 0.9452007	pnorm(1.6, lower.tail = FALSE) 0.05479929

$$P(\bar{\mathbf{S}} < 26 | \bar{\mathbf{S}} \leq 29) = \frac{P(\bar{\mathbf{S}} < 26)}{P(\bar{\mathbf{S}} \leq 29)} = \frac{P\left(Z < \frac{26-28}{5/8}\right)}{P\left(Z < \frac{29-28}{5/8}\right)}$$

$$= \frac{P(Z < -3.2)}{P(Z < 1.6)} = \frac{0.0006871379}{0.9452007} = 0.0007$$

c) What is the **mean length value** that separates **the top 1%** of the **mean lengths** of a **sample** of **64 adult salmon**? Show your work.

You may use the following R output in your calculations.

qnorm(0.01/2, lower.tail = FALSE) 2.575829	qnorm(0.01, lower.tail = FALSE) 2.326348
qnorm(0.05/2, lower.tail = FALSE) 1.959964	qnorm(0.05, lower.tail = FALSE) 1.644854

Top 1% standard normal quantile →

qnorm(0.01, lower.tail = FALSE)

$$z^* = 2.326348$$

Move to the sampling distribution for the sample mean of adult salmon for samples of size 64.

$$\bar{s} = \mu_{\bar{S}} + z^*\sigma_{\bar{S}} = 28 + 2.326348 \times \frac{5}{8} = 29.454$$