MTH 3240 Lab 3

☺ ☺ ☺ Due Thu., Feb. 13 ☺ ☺ ☺

1   Part A: One-Sample t Test (P-Value Approach)

1.1   Radioactivity Data Set

Between 1948 and 1958, the U.S. conducted 43 nuclear tests at Enewetak Atoll in the Pacific Ocean. Cleanup and restoration of the site for the return of the atoll to the Enewetak people began in 1977 and lasted to 1980.

Concern has been expressed by the people of Enewetak and others over the possible aquatic impacts from the radioactive contaminants.

The following data are measured values of the radioactive contaminant radiocesium (^{137}Cs, in pCi/L) in 10 water specimens taken from the site in 1981.

$$69, \quad 77, \quad 227, \quad 215, \quad 28, \quad 153, \quad 25, \quad 325, \quad 55, \quad 42$$

In this problem, we'll carry out a **one-sample t test** to decide if the data provide statistically significant evidence that the water's **population mean radiocesium μ** concentration is **less than 215** pCi/L.

The hypotheses are

$$H_0 : \mu \geq 215$$
$$H_a : \mu < 215$$

where μ is the true (unknown) **population mean radiocesium** concentration, and the **test statistic** is

$$t = \frac{\bar{X} - 215}{S_{\bar{X}}} \qquad \text{where} \qquad S_{\bar{X}} = \frac{S}{\sqrt{n}}.$$

For the **p-value approach**, the **decision rule**, using level of significance $\alpha = 0.05$, is:

Reject H_0 if p-value < 0.05
Fail to reject H_0 if p-value ≥ 0.05

1. Use `mean()` and `sd()` to compute the **sample mean $\bar{X}$** and **sample standard deviation** S of the **radiocesium** data.

2. Use the value of S from Step 1 and the `sqrt()` function to compute the **standard error** $S_{\bar{X}}$ of $\bar{X}$.

3. Use the values obtained in Steps 1 and 2 to compute the **test statistic** t.

4. The function `t.test()` takes as its main argument a vector `x` and carries out a ***one-sample t test*** for a population mean μ and reports the test statistic and p-value. But `t.test()` accepts other optional arguments too. Among its arguments are:

`x`	a data vector.
`alternative`	the direction for the alternative hypothesis, one of `"two.sided"`, `"less"`, or `"greater"`.
`mu`	the null hypothesized value for the unknown population mean, with default value 0.
`conf.level`	the confidence level for a confidence interval for the unknown population mean, with default value 0.95.

We specify the null-hypothesized value μ_0 (**215** in our case) via the argument `mu`. By default, `t.test()` carries out a **two-tailed** test. To carry out a **one-tailed test**, we specify `"less"` or `"greater"` for the optional argument `alternative`. (In this case, the reported confidence interval is one-sided.)

Use `t.test()` to test the hypotheses stated above. For example, if you named your data set `cs`, you'd type:

```
> t.test(cs, mu = 215, alternative = "less")
```

2 Part B: One-Sample t Test (Rejection Region Approach)

2.1 Radioactivity Data Set (Cont'd)

We want to repeat the one-sample t test of

$$H_0 : \mu \geq 215$$
$$H_a : \mu < 215$$

but this time using the ***rejection region approach***.

In this case the ***decision rule***, using level of significance $\alpha = 0.05$, is:

Reject H_0 if $t < -t_{0.05,n-1}$
Fail to reject H_0 if $t \geq -t_{0.05,n-1}$

where the ***t critical value $t_{0.05,n-1}$*** is 95th percentile (or *quantile*) of the t distribution with $n-1$ degrees of freedom.

1. The function `qt()` will compute ***t critical values***. It takes arguments:

`p`	a lower tail probability (e.g p = 0.95 would correspond to a 95th percentile).
`df`	the degrees of freedom (e.g. the value of $n-1$).

and returns the corresponding ***t critical value***.

Use `qt()`, with `p = 0.95` and `df = 9`, to compute the ***t critical value $t_{0.05,n-1}$***.

Compare your **test statistic value** t from Part A to to $-t_{0.05,9}$.

3 Part C: One-Sample t Test on Log-Transformed Data

3.1 Radioactivity Data Set (Cont'd)

In Part A, you carried out a one-sample t test to decide if the true (unknown) population mean radiocesium concentration in the water is less than 215 pCi/L.
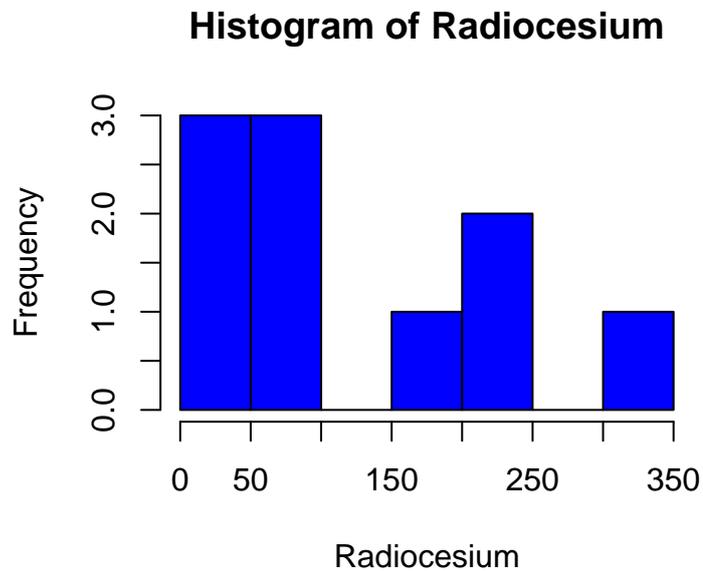
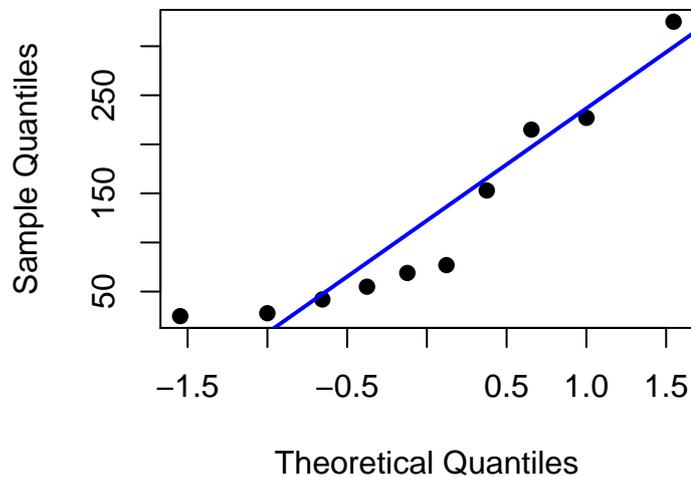The hypotheses were

$$H_0 : \mu \geq 215$$
$$H_a : \mu < 215$$

1. The **one-sample t test** rests on the assumption that either the sample was drawn from a **normal population** or n **is large**. If neither condition is met, the p-value from the t test might be wrong.

   We can check the normality assumption using the plots below.

Histogram of Radiocesium

Normal Q–Q Plot

Both plots hint that the sample is from a **right-skewed** population, so the normality assumption is questionable.

2. Take the **logs** of the radiocesium data by typing:

```
logcs <- log(cs)
```

(assuming your original vector of radiocesium data is named `cs`).

3. Use

```
hist(logcs, main = "Histogram of Log Radiocesium",
     col = "blue", xlab = "Log Radiocesium")
```

and

```
qqnorm(logcs, pch = 19)
qqline(logcs, col = "blue", lwd = 2)
```

to check the normality assumption for the **logs** of the radiocesium data.

4. The normality assumption should be met (approximately) for the **logs** of the radiocesium data, and the t test can be carried out on the **logs**.

Because we're working on the log scale, we'll test the hypotheses

$$H_0 : \mu_{\log} \geq 5.37$$
$$H_a : \mu_{\log} < 5.37,$$

4

where $\mu_{\log}$ is the (unknown) **population mean *log* radiocesium concentration** and the value **5.37** was obtained by computing

```r
log(215)
```

Use **t.test()** to carry out ***one-sample t test*** of the above hypotheses using the **logs** of radiocesium data. For example, if you named your data set `logcs`, you'd type:

```r
t.test(logcs, mu = 5.37, alternative = "less")
```