Southern University of Science And Technology
School of Life Sciences
**BIO210 Biostatistics**
(Spring, 2024)

# Assignment 5
# Due on 28th April, 11 p.m.

**1.** **(7.5 points)** **One-sided confidence interval 1**: A random sample of size $n$ has been drawn from a population that follows a normal distribution with the standard deviation $\sigma$. Follow similar procedures described in **Lecture 19**, construct two ***one-sided*** 95% confidence intervals for the population mean, one for a lower bound and one for an upper bound.

**2.** **(7.5 points)** **One-sided confidence interval 2**: A random sample of size $n$ has been drawn from a population that follows a normal distribution with unknown standard deviation. Construct two ***one-sided*** 95% confidence intervals for the population mean, one for a lower bound and one for an upper bound. **Note:** the value of $\boldsymbol{t}_{\alpha,\nu}$ represents the $t$ score with the degree of freedom $\nu$ where the probability of the upper tail is $\alpha$.

**3.** A random sample of 500 sales prices of recently purchased homes in a county is taken. From that sample a 90% confidence interval for the average sales price of all homes in the county is computed to be £$215,000 \pm 35,000$. Is the following statement true or false? Put **T** (true) or **F** (false) in the box at the front.

[ ] **(2 point)** There is a 90% chance that the average sales price of all homes in the county is in the range £$215,000 \pm 35,000$.

[ ] **(2 point)** About 90% of all home sales in the county have a sales price in the range £$215,000 \pm 35,000$.

A poll of 400 eligible voters in a city finds that 313 plan to vote in the next election. The 95% confidence interval for the percentage of all eligible voters in the city who plan to vote is **(2 point)** ____________. Denote this interval as $[a, b]$. Is the following statement true or false? Put **T** (true) or **F** (false) in the box at the front.

[ ] **(2 point)** In repeated sampling, there is a 95% chance that a sample will produce the same interval as $[a, b]$.

[ ] **(2 point)** In repeated sampling, about 95% of samples will produce intervals that cover the population proportion.

**4.** Write out the null ($H_0$) and the alternative ($H_1$) hypotheses for the following scenarios:

**4.1)** **(5 points)** An engineering department had 7% of female students. The department had been working hard to increase the percentage of female students. A researcher wanted to test if the female student rate had actually increased, so he obtained a random sample of students to see what proportion of the sample was female.

$\boldsymbol{H_0}$:

$\boldsymbol{H_1}$:

**4.2)** **(5 points)** Li Lei is a college basketball player who has scored 75% of the free-throws he has attempted in his career. He decided to practice a new technique for shooting his free-throws. Li Lei was curious if this new technique produced significantly better or worse results. He tried the new technique and made 70% of 50 attempts.

$\boldsymbol{H_0}$:

$\boldsymbol{H_1}$:

**4.3)** **(5 points)** A restaurant owner installed a new automated drink machine. The machine is designed to dispense 530 mL of liquid on the medium size setting. The owner suspects that the machine may be dispensing too much in medium drinks. They decide to take a sample of 30 medium drinks to see if the average amount is significantly greater than 530 mL.

$\boldsymbol{H_0}$:

$\boldsymbol{H_1}$:

**4.4)** **(5 points)** A healthcare provider saw that 48% of their members received their flu shot in a recent year. The healthcare provider tried a new advertising strategy in the following year, and they took a sample of members to test if the proportion who received their flu shot had changed.

$\boldsymbol{H_0}$:

$\boldsymbol{H_1}$:

**5.** **(5 point)** **The *p*-value**: Is the following statement true or false about the $p$-value? Put **T** (true) or **F** (false) in the box at the front.

[ ] A $p$-value of 0.001 indicates the null hypothesis is wrong.

[ ] If the $p$-value is 0.001, then there the probability that the null hypothesis is true is only 0.001, so we should reject the null hypothesis.

[ ] If the null hypothesis is true, then there is less than a 5% chance to get a $p$-value that is smaller than 5%.

[ ] If we get a $p$-value of 0.2, it means the probability of the null hypothesis being true is 0.2, which is not small, so we should not reject the null hypothesis

[ ] If we get a $p$-value of 0.02, it means the probability of the alternative hypothesis being wrong is only 0.02, which is really small, so we should reject the null hypothesis and accept the alternative hypothesis

**6. Blue M&M's candies**: M&M's® have been around since 1941. The "plain" M&M's candies (now called "milk chocolate M&M") are produced by the Mars, Inc. company. There are six colours of the plain M&M's candies: blue, brown, green, orange, red & yellow. The distribution of colours in M&M's has a long and colourful history. The colours and proportions occasionally change, and the distribution is different for different types of candies. In 2008, the company changed to the following distribution:

| Colour | blue | brown | green | orange | red | yellow |
|---|---|---|---|---|---|---|
| % | 24 | 13 | 16 | 20 | 13 | 14 |

In 2017, Li Lei wanted to test whether plain M&M's candies really contain 24% blue ones as claimed back in 2008. He obtained a simple random sample containing 500 plain M&M's candies and the proportion of blue candies was 18.4%.

**6.1)** **(2.5 points)** The population represented by those 500 candies is:

(A) All M&M's candies.

(B) All plain M&M's candies in 2017

(C) All plain M&M's candies produced since 1941

(D) All plain M&M's candies that have been produced and will be produced using the same colour distributions used in 2017

**6.2)** **(2.5 points)** Write the null and alternative hypotheses for the test.

**6.3)** **(2.5 points)** Compute the test statistic.

**6.4)** **(2.5 points)** If $H_0$ were true, what will be the distribution of the test statistic?

**6.5)** **(2.5 points)** Apparently, 18.4% $\neq$ 24%. What might be the reasons causing the difference?

**6.6)** **(2.5 points)** What is the two-sided 95% confidence interval for the true population proportion based on the data from those 500 candies?

**6.7)** **(2.5 points)** What is the $p-$value of the test, and should Li Lei reject $H_0$ at the significance level of 0.01?

**7. Renal Disease**: The mean serum-creatinine level measured in 12 patients one day after they received a newly proposed antibiotic was 1.2 mg/dL. Suppose the mean and standard deviation of serum creatinine in the general population are 1.0 and 0.4 mg/dL, respectively, and suppose the new antibiotic does not change the dispersion of the serum-creatinine level. Use a significance level of 0.05, perform a hypothesis testing to see whether the mean serum-creatinine level in this group is different from that of the general population.

**7.1)** **(2.5 points)** The population represented by these 12 patients is:

(A) All people.

(B) All healthy people

(C) All patients that have taken the new antibiotic

(D) All patients that have NOT taken the new antibiotic

**7.2)** **(2.5 points)** Write the null and alternative hypotheses for the test.

**7.3)** **(2.5 points)** Compute the test statistic.

**7.4)** **(2.5 points)** If $H_0$ were true, what will be the distribution of the test statistic?

**7.5)** **(2.5 points)** What is the two-sided 95% confidence interval for the patient population mean based on the data from the 12 patients.

**7.6)** **(2.5 points)** What is the $p-$value of the test, and do you reject $H_0$ ?

**8. Lady Tasting Tea**: British Tea is quite different from Chinese tea. The addition of milk to tea has been a cultural tradition in Britain for centuries. The practice of adding milk to tea has its roots in the history of tea consumption and the preferences of the British people. Different people have different habits. Some put tea first into the cup, but some put milk first. One of the primary reasons for adding milk to tea in Britain is to mitigate the strong and sometimes bitter flavour of black tea. The famous "**Lady tasting tea**" problem is a well-known statistical problem that was posed by Sir Ronald A. Fisher. This problem played a significant role in the development of modern statistics and the establishment of Fisher as one of the pioneers of the field. The scenario is like this:

At a tea party, a lady claims she can discriminate by tasting a cup of tea whether the milk or the tea was poured in the cup first. To verify her claim, a statistician presents her in random order with eight cups of tea, and tells her there are exactly four cups of tea in which tea is poured in first. To make the situation easy to describe, we denote those cups of tea where the milk is poured in first as type **M**, and those where the tea is poured in first as type **T**. Apparently, we have four **M**s and four **T**s. How can the statistician verify if the lady's claim is true or not? Let's work this out in a step-by-step manner.

**8.1)** **(5 points)** Apparently, even if the lady randomly assigns **M** or **T** to those cups of tea, there is still a chance that she gets some of them correct. We want to test if the lady has the ability to tell the correct types of those cups of tea. Write the null and the alternative hypotheses.

**8.2)** **(7.5 points)** Under the null hypothesis, we define a random variable $\boldsymbol{X}$ to represent the number of cups of tea to which the lady correctly assigns the type. Compute the PMF of $\boldsymbol{X}$ and $\mathbb{E}[\boldsymbol{X}]$. Explain the meaning of $\mathbb{E}[\boldsymbol{X}]$.

**8.3)** **(5 points)** Eventually, it turns out that the lady correctly identifies the four **M** types (and thus also correctly identifies the four **T** types). What is the $p-$value in this case?