UNIVERSITY OF TEXAS AT AUSTIN, DEPARTMENT OF MATHEMATICS
M358K - Applied Statistics

HOMEWORK #4

Provide your *final answer only* to the following problems:

Problem 4.1. (5 points)
In a hypothesis testing problem, $p-$value $= 3\%$ means that ...

a.: Null hypothesis has a 3% chance to be wrong.
b.: If the null hypothesis is true, the probability of observing as extreme or more extreme than what have been observed is 3%.
c.: Alternative hypothesis has a 3% chance to be wrong.
d.: If we repeat the procedure a lot times, approximately 3% of the tests will be significant.
e.: None of the above.

Solution: b.

Provide your **complete solution** to the following problems. Even if correct, the final answer without the correct justification will earn zero points.

Problem 4.2. $(5 + 3 + 2 = 10$ pts) *Source: "Probability and Statistics for Engineers and Scientists" by Walpole, Myers, Myers, and Ye.*
The average zinc concentration recovered from a sample of measurements taken at 36 different locations in a river is found to be 2.6 grams per milliliter. Assume the normal distribution for the concentration and let the population standard deviation be 0.3 gram per milliliter. Find the

(i) 95% confidence interval, and
(ii) 99% confidence interval

for the mean zinc concentration in the river. Moreover, how large a sample is required if we want to be 95% confident that our estimate of the mean parameter is off by less than 0.05?

Solution: The point estimate of the population mean is $\bar{x} = 2.6$. The standard error is $\frac{\sigma}{\sqrt{n}} = \frac{0.3}{\sqrt{36}} = 0.05$.

(i) For a 95% confidence interval, the critical value is $z^* = \Phi^{-1}(0.025) = 1.96$. So, the 95%−confidence interval is

$$2.6 - (1.96)(0.05) < \mu < 2.6 + (1.96)(0.05),$$

i.e., 2.6 ± 0.098, i.e., $(2.502, 2.698)$.

(ii) For a 99% confidence interval, the critical value is $z^* = \Phi^{-1}(0.005) = 2.575$. So, the 99%−confidence interval is

$$2.6 - (2.575)(0.05) < \mu < 2.6 + (2.575)(0.05),$$

i.e., 2.6 ± 0.12875, i.e., $(2.47125, 2.72875)$.

As for the required sample size so that we are 95% confident that our estimate of the mean parameter is off by less than 0.05, the condition is

$$(1.96)\left(\frac{0.3}{\sqrt{n}}\right) \leq 0.05 \quad \Rightarrow \quad n \geq \left(\frac{1.96(0.3)}{0.05}\right)^2 = 138.2976.$$

Since n must be an integer, the condition is $n \geq 139$.

Problem 4.3. (5 points) *Source: "Mathematical Statistics with Applications" by Ramachandran and Tsokos.*
Fifteen vehicles were observed at random for their speeds (in mph) on a highway with a speed limit posted as 70 mph. It was found that the observed sample average was 73.3 mph. Suppose that from the past experience we can assume that vehicle speeds are normally distributed with the standard deviation of the speed equal to 3.2 mph. Construct a 90% confidence interval for the true mean speed μ of the vehicles on this highway.

Solution: see p. 302.

Problem 4.4. (15 points) Suppose that the thumb sizes of the US males follow a normal distribution with an unknown mean μ and known standard deviation $\sigma = 20$ on the GPI - scale (*Grey's Pollex Index - GPI - from* 50 *to* 280). The US Department of Thumbs and Toes (DTT) reports that the mean thumb size in the country is $\mu = 150$.

Being the chairman of the Faculty of Thumbs of the local university you see an excellent opportunity here and decide to conduct your own study of the size of the average American thumb. After collecting a SRS of 100 American thumbs you obtain the following sample average $\bar{x} = 153$.

i (5 pts) Construct a 95%-confidence interval for the unknown parameter μ based on your study.
Solution:

$$\bar{x} \pm \frac{z^*\sigma}{\sqrt{n}} \text{ i.e. } 153 \pm 3.92 \text{ i.e. } (149.08, 156.92).$$

ii. (8 pts)
Assess the strength of evidence your study carries against the DTT findings. In other words: state the hypotheses and report the $p-$value.
Solution:
(2 points) The hypotheses are

$$\begin{cases} H_0: & \mu = 150 \\ H_a: & \mu \neq 150 \end{cases}$$

(6 points) To get the p-value we calculate $2\mathbb{P}[\bar{X} > 153]$. A simple z-score calculation gives us that the p-value is $2(0.0668) = 0.1336$.

(iii) (2 pts) You dream of achieving fame and fortune by being the first person ever to estimate the mean thumb size up to the margin of error equal to ± 0.1. How large a sample size do you need for that?
Solution:

$$n \geq \left(\frac{z^*\sigma}{0.1}\right)^2 = (1.96(200))^2 = 153664.$$

Problem 4.5. (10 points)
Source: Ramachandran, Tsokos.

It is claimed that sports-car owners drive on the average $20,000$ miles per year. A consumer firm believes that the mean annual mileage is actually <u>lower</u>. To check, the consumer firm decided to test this hypothesis.

The modeling assumptions are that the annual mileage is normally distributed with an unknown mean μ and with the standard deviation of 1200.

The consumer firm obtained information from 36 randomly selected sports-car owners that resulted in a sample average of $19,530$ miles. What is the decision of this hypothesis test at the signficance level of 0.01?

Solution:
The null and the alternative hypotheses are

$$H_0 : \mu = 20000 \quad vs. \quad H_a : \mu < 20000.$$

The observed value of the $z-$statistic is (under the null hypothesis), in our usual notation,

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{19530 - 20000}{1200/\sqrt{36}} = -2.35.$$

On the other hand, the $z-$score corresponding to the left-sided test at the 0.01 significance level is -2.33. Because the observed value of the z statistic -2.35 is less than -2.33, the null hypothesis is **rejected** at the significance level of 0.01. There is sufficient evidence to conclude that the mean mileage on sport cars is less than 20,000 miles per year.

Problem 4.6. (5 points) *Source: "Probability and Statistics for Engineers and Scientists" by Walpole, Myers, Myers, and Ye.*
Assume that the compressive strength for a certain type of cement is normal with a known standard deviation of 120 kilograms and an **unknown** mean μ. You test the hypotheses

$$H_0 : \mu = 5000 \quad vs. \quad H_a : \mu < 5000.$$

For a planned sample of size 50, your colleague obtains the rejection region $RR = (-\infty, 4970]$. What is the significance level he used?

Solution: For the left-sided alternative, the upper bound rejection region is of the form

$$\mu_0 + z_\alpha \left(\frac{\sigma}{\sqrt{n}}\right).$$

So, in this problem, we have that

$$5000 + z_\alpha \left(\frac{120}{\sqrt{50}}\right) = 4970 \quad \Rightarrow \quad z_\alpha = \frac{4970 - 5000}{\frac{120}{\sqrt{50}}} = -1.767767.$$

We can use **R** or the standard normal tables at this point. I decided to use **R**. The command gives me $\alpha = 0.0385$.