

Q. 1 (Usual hypothesis testing with one-sided alternative:). The managers at Hunter Chemical Company claim that their major product contains on average $\mu = 4.0$ fluid ounces of caustic material per gallon. They further state that the distribution of caustic material is normal with a standard deviation of $\sigma = 1.3$ fluid ounces. If there is too much caustic material, the product will be dangerous. A government inspector is brought in to test the product. She randomly selects a sample of 100, gallon-size containers of the product and finds that their mean weight of caustic materials is $\bar{X} = 4.5$ fluid ounces per gallon.

- (a) What are the null and alternative hypotheses that the inspector is exploring?
Our default position is that the company is telling the truth. Our alternative is that something is wrong and that the product is dangerous. $H_0 : \mu = 4$
 $H_a : \mu > 4$
- (b) Explain in English what the type I and type II errors for this problem would be?
If the product indeed contained 4.0 fluid ounces of caustic material per gallon, but we rejected H_0 , that would be a Type 1 error. On the other hand, if the product did not contain 4.0 fluid ounces of caustic material per gallon, but we did not reject the null hypothesis, i.e., concluded that H_0 was true, that would be a Type II error.
- (c) What is the probability of getting a sample mean as high as or higher than 4.5, if the production process was working as the Company claims? Based on this, state your conclusion of the test. Reject at level 0.05.
Since we know standard deviation and n is large enough, \bar{X} is normal by the CLT and we will use a Z calculation (instead of t). The probability of the inspector seeing a sample mean of 4.5 or more when the process was really producing a mean of 4:

$$P(\bar{X} \geq 4.5 | \mu = 4) = P(Z \geq \frac{4.5 - 4}{1.3/\sqrt{100}}) = P(Z \geq 3.85) = 0.0001.$$

The p-value is 0.0001, so we reject the null hypothesis.

Q. 2 (Find value of statistic for a given level of test). A tutoring program claims to increase IQ of children. They cite a study of 25 children randomly samples from all over Ontario that showed an average IQ of 104 after going through their program. We happen to know that in the general population of children, the average IQ is 100 and the population SD is 15. We want to test whether the program is effective or not.

- (a) What are the null and alternative hypotheses?
Our default position is that the mean IQ remains the same. Our alternative is that the tutoring program indeed increases the IQ of children.
 $H_0 : \mu = 100$
 $H_a : \mu > 100$

- (b) Perform a test based on the sample. Reject at level 0.05.
 p value = $P(Z > \frac{104-100}{15/\sqrt{25}}) = P(Z > 1.3333) = 0.0912$. If the null was true, the probability of obtaining an IQ of 104 or more is 0.0912. > 0.05 . Cannot reject the null.
- (c) How high would the sample mean have to be for you to reject the null hypothesis at the level 5%?
 We want p value = $P(Z > \frac{\bar{X}-100}{15/\sqrt{25}}) \leq 0.05$. From the standard normal distribution, we know that $P(Z > 1.645) = 0.05$. Therefore $\frac{\bar{X}-100}{15/\sqrt{25}} \geq 1.645$. Solving gives, $\bar{X} \geq 104.935$.

Q. 3 (CI and hypothesis testing). The MSCEIT is a measure of Emotional Intelligence. It has a mean of 100 and a standard deviation of 15 in the population. A sample of 49 people was collected. The sample mean was 99.

- (a) Calculate a 90% confidence interval for the population mean.
 $\hat{\sigma}_{\bar{X}} = 15/\sqrt{49} = 2.1429$. Since the population variance is known, we use a z-statistic. The interval is therefore, $[99 - 1.645 * 2.1429, 99 + 1.645 * 2.1429] = [95.45, 102.52]$.
- (b) Only based on your answer to part (a), if you had tested the null hypothesis that the population mean was 100, would you have retained or rejected it? Use the significance level of 10%.
 We know that μ_0 lies in the confidence interval for μ if and only if the hypothesis test accepts the null hypothesis. In other words the confidence interval consists precisely of all those values μ_0 for which the null hypothesis $H_0 : \mu = \mu_0$ is accepted. Since our interval in part (a) contains 100, we accept the null.

Q. 4 (Expectation). Suppose X has the probability density

$$f(x) = \frac{1}{2\sqrt{x}}$$

for $0 < x < 1$.

- (a) For what positive integers X does X^k have expectation?

$E(X^k) = \int_0^1 \frac{x^k}{2\sqrt{x}} = \frac{1}{2(k+1/2)} x^{k+1/2} \Big|_0^1 = \frac{1}{2k+1}$. Thus expected value exists for all positive integers k .

- (b) Calculate $E(X^k)$ for the positive integers k such that expectation exists
 See above

Q. 5 (Expectation). Suppose X has the probability density

$$f(x) = \frac{1}{6}$$

for $0 < x < 2$ and

$$f(x) = \frac{1}{3}$$

for $2 < x < 4$.

(a) Find the median

The median is the value M such that $P(0 < X < M) = 1/2 = P(M < X < 4)$.

Now we do not know if M is greater than or lesser than 2. If M is greater than 2 then we get $\int_0^2 \frac{1}{6} + \int_2^M \frac{1}{3} = \frac{2}{6} + \frac{M-2}{3} = 1/2$ and also $\int_M^4 \frac{1}{3} = \frac{4-M}{3} = 1/2$. Solving, we get $M = 2.5$.

Now just to check, alternatively, if M is lesser than 2, then similarly we get $\frac{2}{3} + \frac{2-M}{6} = 1/2$ and $\frac{M}{6} = 1/2$. Then solving we get $M = 3$ which is not less than 2. Hence the median is 2.5 as found above.

(b) Find the median of the sum of 40 such random variables. (Use CLT)

By the CLT $X_1, X_2, \dots, X_{40} \sim N(40\mu, \sqrt{40}\sigma)$ where μ and σ are the mean and standard deviation of X . We can use usual integration to find that $\mu = \frac{7}{3}$. Now for normal random variables, the mean is equal to the median and hence the median is $40 * \frac{7}{3} = 93.33$.

Q. 6 (CI). How many questions need to be on a short answer test for us to be 95% sure that someone who knows 80% of the course material gets a grade between 75% and 85%?

The expected value of the proportion of correct answers is 0.8 and the variance for the proportion is $0.8 * 0.2/n = 0.16/n$. The 95% confidence interval for the proportion of correct questions is then $0.8 \pm 1.96 * \sqrt{0.16/n}$. Now since we need the score to be between 75% and 85%, we want to be 0.5% away from the mean on each side. Hence $1.96 * \sqrt{0.16/n} = 0.05$ and solving we get $n \geq 246$.

Q. 7 (Probability). W, X, Y, Z independent standard normal. What is the probability that $W + X > Y + Z + 1$?

We could do this with integrals but there is a much simpler way.

Note our condition is equivalent to $W + X - Y - Z > 1$, and $-Y, -Z$ are also standard normals. Now $W + X - Y - Z \sim N(0, 4)$, and thus we are asking what is the probability that a normal random variable with mean 0 and variance 4 is at least 1. Let us set $S = (W + X - Y - Z)/2$. Note the mean of S is 0 and the variance is 1, hence it is standard normal! Our question now becomes what is $P(S > 1/2)$ where $S \sim N(0, 1)$.

Thus the answer is $1 - \phi(.5)$ where $\phi(\cdot)$ is the standard normal cdf.