Math 10 - Practice Problems for Final Exam

Brand New Sample Exam Problems

Sample Exam Question 1 - 15 points

The monthly sales of a company's product (Y, in thousands of units) is linearly related with the monthly advertisement spending (X, in thousands of dollars). You want to model this using linear regression with dependent variable Y and independent variable X. Data for the last $n = 7$ months and their summary statistics are shown below.

X	3	5	8	2	5	6	10
Y	5	11	14	4	11	11	19

Sample means: $\bar{X} = 5$, $\bar{Y} = 10.7$.

Sample standard deviations (estimators of the standard deviation) : $s_X = 6.6$, $s_Y = 26.4$.

Pearson Correlation $r = 0.95$.

a) Calculate the slope and intercept cofficient, then write down the regression line. *(5 points)*

b) Using only this regression line, predict the value of Y when $X = 6$. *(2 points)*

c) Using only this regression line, predict the change in the monthly sales of the product when advertisement spending is increased by 2 thousand dollars. *(2 points)*

d) You are given that the standard error of the slope is $s_b = 2$. Perform a two-tailed test on whether the slope coefficient b is significantly different from zero, at a significance level of $\alpha = 0.05$, and write a conclusion. *(6 points)*

Answer

a) $b = r \cdot \frac{s_Y}{s_X} = 0.95 \cdot \frac{26.4}{6.6} = 0.95 \cdot 4 = 3.8$. And $a = 10.7 - 3.8 \cdot 5 = 10.7 - 19 = -8.3$. Then, regression line is $\hat{Y} = 3.8X - 8.3$.

b) The predicted value is $\hat{Y} = 3.8 \cdot 6 - 8.3 = 22.8 - 8.3 = 14.5$.

c) Slope is $b = 3.8$. So, the predicted change in Y is an increase of $2 \cdot 3.8 = 7.6$ thousand units of product.

d) $H_0 : \beta = 0$, $H_A : \beta \neq 0$.

Degrees of freedom $df = n - 2 = 7 - 2 = 5$.

$P(\textit{ sample slope } \geq 3.8) = P(T \geq \frac{3.8}{2}) = P(T \geq 1.9) > P(T \geq 2.57) = \frac{\alpha}{2} = 0.025$.

Condition for rejecting the null is not met. So, we do not reject the null at $\alpha = 0.05$ significance level. Hypothesis test is inconclusive.

Sample Exam Question 2 - 15 points

The biological trait Y of offsprings are modeled with linear regression to be dependent on the biological trait X of parents. The sample data used in this model are as follows.

Offspring trait, X	5	2	3	8	9	7
Parent trait, Y	4	4	2	10	11	9

Sample means: $\bar{X} = 6$, $\bar{Y} = 6.7$.

Sample standard deviations (estimators of the standard deviation) : $s_X = 7.1$, $s_Y = 14.2$.

Pearson Correlation $r = 0.80$.

a) Calculate the slope and intercept cofficient, then write down the regression line. *(5 points)*

b) You are given that the standard error of the slope is $s_b = 0.5$. Perform a **one-tailed test** on whether the slope coefficient b is significantly **greater** than zero, at the $\alpha = 0.03$ level of significance. Write a conclusion. *(6 points)*

c) Using only this regression line, predict the value of Y when $X = 10$. *(2 points)*

d) Using only this regression line, predict the change in trait Y when trait X is decreased by 5 units. *(2 points)*

Answer

a) $b = r \cdot \frac{s_Y}{s_X} = 0.80 \cdot \frac{14.2}{7.1} = 0.80 \cdot 2 = 1.6$. And $a = 6.7 - 1.6 \cdot 6 = 6.7 - 9.6 = -2.9$. Then, regression line is $\hat{Y} = 1.6X - 2.9$.

b) $H_0 : \beta = 0$, $H_A : \beta > 0$.

Degrees of freedom $df = n - 2 = 6 - 2 = 4$.

$P(\text{ sample slope } \geq 1.6) = P(T \geq \frac{1.6}{0.5}) = P(T \geq 3.2) < P(T \geq 2.78) = 0.025 < 0.03$.

Condition for rejecting the null has been met. So, we reject the null hypothesis at $\alpha = 0.03$ significance level. The slope coefficient is probably greater than zero.

c) The predicted value is $\hat{Y} = 1.6 \cdot 10 - 2.9 = 16 - 2.9 = 13.1$.

d) Slope is $b = 1.6$. So, the predicted change in Y is $1.6 \cdot (-5) = -8$ (or a decrease of 8 units in Y).

Sample Exam Question 3 - 5 points

There are equal proportions of cards of 3 colors Red, Blue and Green in a large deck of cards. You shuffled the deck of cards to the best of your ability, then took 12 cards from the top. If the shuffling was done perfectly, you would **expect** an equal number of Red, Blue and Green cards in those 12.

Looking at the 12 cards you took, you **observe** that 6 are Red, 2 are Blue and 4 are Green. Perform a hypothesis test at the $\alpha = 0.10$ level of significance on whether the observed data differs significantly from the expected values.

Write a conclusion: based on this hypothesis testing alone, was your shuffling significantly flawed? *(5 points)*

Answer

Expected frequencies : $4, 4, 4$
Observed frequencies : $6, 2, 4$ (Red, Blue, Green)

H_0 : distributions same.
H_A : distributions different.

Degrees of freedom $= 3 - 1 = 2$.

$\chi^2_{df=2} = \frac{(4-6)^2}{4} + \frac{(4-2)^2}{4} + \frac{(4-4)^2}{4} = \frac{(-2)^2}{4} + \frac{2^2}{4} = 2.$

$P(\chi^2_{df=2} \geq 2) > P(\chi^2_{df=2} \geq 4.61) = 0.10$. So, condition for rejecting the null is not met. We do not reject the null, at $\alpha = 0.10$ significance level. The hypothesis test is inconclusive. Optional additional conclusion: the shuffling was probably not significantly flawed.

Sample Exam Question 4 - 10 points

A company has a set of old data showing the percentages of customers and their preference for product A,B or C. This set of data showed that 25% preferred product A, 25% preferred B, and 50% preferred C.

The company acquired a new set of data consisting of $n = 100$ new customers and their preferences. In this new set of data, you **observed** that product preferences are: $A = 20$, $B = 20$, $C = 60$.

You want to do a Chi Square hypothesis test on whether the numbers **observed** in this new set of data differs from the numbers **expected** from the old data.

a) Write down the null and alternative hypothesis. *(2 points)*

b) Calculate the test statistic and state the degrees of freedom. Hint: you need to convert the percentages to numbers. *(5 points)*

c) Perform the Chi Square hypothesis test at $\alpha = 0.20$ significance level. Write a conclusion. *(3 points)*

Answer

a) H_0 : distributions same. H_A : distributions different.

b) Expected frequencies : $25, 25, 50$. Observed frequencies : $20, 20, 60$ (A, B, C)

Degrees of freedom $= 3 - 1 = 2$.

$\chi^2_{df=2} = \frac{(25-20)^2}{25} + \frac{(25-20)^2}{25} + \frac{(50-60)^2}{50} = 2 \cdot \frac{5^2}{25} + \frac{10^2}{50} = 2 + 2 = 4.$

c) $P(\chi^2_{df=2} \geq 4) < P(\chi^2_{df=2} \geq 3.22) = 0.20.$

Condition for rejecting the null has been met. We reject the null, at $\alpha = 0.20$ significance level. Conclusion: the new set of data has a distribution that probably differs from the old one.

Sample Exam Question 5 - 10 points

You have $k = 3$ samples of size $n = 7$ each. Each of these samples are drawn from a normal distribution with the same variances but possibly different respective means μ_1, μ_2, μ_3.

Sample variances : $s_1^2 = 12, s_2^2 = 17, s_3^2 = 16$. (estimators of the variance)

The (sample) variance of the sample means is 8.6.

Write down the null and alternative hypothesis, then perform an ANOVA hypothesis test at $\alpha = 0.05$. Write a conclusion. *(10 points)*

Answers

Null: $\mu_1 = \mu_2 = \mu_3$, alternative: some means are different.

$MSB = 7 \cdot 8.6 = 60.2$.
Degrees of freedom: $df_1 = k - 1 = 3 - 1 = 2$.

$MSE = \frac{12+17+16}{3} = \frac{45}{3} = 15$.
Degrees of freedom: $df_2 = N - k = 21 - 3 = 18$.
w $F = MSB/MSE = 60.2/15 = 4.01$ (you can round numbers to nearest 2 decimal places in the exam).

F statistic has degrees of freedom: $df_1 = 2$ and $df_2 = 18$.

$P(F \geq \frac{60.2}{15}) = P(F \geq 4.01) < P(F \geq 3.55) = 0.05 = \alpha$. Alternatively, you could say that the F-statistic is greater than 4.01, which is greater than the critical F-value of 3.55. Condition has been met, so the null is rejected at 0.05 level of significance. Some means are probably different.

Sample Exam Question 6 - 10 points

You want to figure out if the two different schools produces students with different scores on a test on average, or is one producing students with higher score on average. You took two samples, both of size 10, of students from each school. Summary statistics from your samples are:

Sample 1 from school 1 : sample mean $\bar{X}_1 = 50$ points on the test, estimator of variance $s_1^2 = 90$.
Sample 2 from school 2 : sample mean $\bar{X}_2 = 60$ points on the test, estimator of variance $s_2^2 = 70$.

Suppose that scores from both schools are normally distributed, with unknown variances. Let μ_1 and μ_2 be the respective school's population means.

Perform a one-tailed hypothesis test on whether the means of both schools are equal, or if school 2 has a higher mean test score than school 1, at $\alpha = 0.025$ significance level. Write a conclusion. *(10 points)*

Answers

$H_0 : \mu_2 - \mu_1 = 0$.
$H_A : \mu_2 - \mu_1 > 0$.

Degrees of freedom, $df = (10 - 1) + (10 - 1) = 18$.

Standard error, SE $= \sqrt{\frac{s_1^2 + s_2^2}{n}} = \sqrt{\frac{160}{10}} = \sqrt{16} = 4$.

Sample difference $= 60 - 50 = 10$. **NOTE:** The question asks for population 2 minus population 1.
$P(\textit{ sample difference } \geq 10) = P(T \geq \frac{10}{4}) = P(T \geq 2.5) < P(T \geq 2.10) = 0.025$.

Condition for rejecting the null has been met. So, we reject the null hypothesis at $\alpha = 0.025$ significance level. School 2 probably has a higher mean test score than school 1.

Sample Exam Question 7 - 15 points

a) X is a variable from a population which has a distribution with unknown mean μ and known variance $\sigma^2 = 16$. You take a sample $\{X_1, X_2, ..., X_{16}\}$ of size $n = 16$ from this population and calculated a sample mean of $\bar{X} = 9.5$.

Perform a hypothesis test on whether the mean $\mu = 11$ or if the mean $\mu < 11$ at $\alpha = 0.10$ significance level. Write a conclusion. *(5 points)*

b) Y is a variable from a population, with a normal distribution with unknown mean μ and unknown variance. You take a sample $\{Y_1, Y_2, ..., Y_9\}$ of size $n = 9$ from this population and calculated a sample mean of $\bar{Y} = 11$, and a sample variance of $s^2 = (1.5)^2$ (estimator of the variance).

Perform a two-tailed hypothesis test on whether the mean $\mu = 10$ or if the mean $\mu \neq 10$, at the $\alpha = 0.05$ level of significance. *(5 points)*

c) A Christmas display contains a large amount of red and blue balls. You want to know if the manufacturer supplied 70% blue balls and 30% red balls as ordered. You took a sample of $n = 21$ balls and calculated a sample proportion of $p = 0.60$ blue balls (just pretend that this sample proportion works :p).

Perform a hypothesis test on whether the population proportion of blue balls is 0.70, or if it is less than what you ordered, at $\alpha = 0.10$ significance level. Write a conclusion. *(5 points)*

Answers

a) $H_0 : \mu = 11$, $H_A : \mu < 11$. Standard error, $\text{SE} = \frac{\sigma}{\sqrt{n}} = \frac{4}{4} = 1$. $P(\textit{ sample mean } \leq 9.5) = P(Z \leq -1.5) = 0.0668 < 0.10$. Condition has been met. We reject the null at $\alpha = 0.10$ significance level. Population mean is probably less than 11.

b) $H_0 : \mu = 10$, $H_A : \mu \neq 10$. Degrees of freedom $df = 9 - 1 = 8$. Standard error, $\text{SE} = \frac{s}{\sqrt{n}} = \frac{1.5}{3} = 0.5$. $P(\textit{ sample mean } \geq 11) = P(T \geq \frac{11-10}{0.5}) = P(T \geq 2) > P(T \geq 2.31) = 0.025$. Condition is not met. We do not reject the null at $\alpha = 0.05$ significance level. Hypothesis test is inconclusive. Optional additional conclusion: the population mean is probably not significantly different from 10.

c) $H_0 : \pi = 0.70$, $H_A : \pi < 0.70$. Standard error $\text{SE} = \sqrt{\frac{0.7 \cdot 0.3}{21}} = \sqrt{\frac{0.21}{21}} = \sqrt{0.01} = 0.1$, $P(\textit{ sample proportion } \leq 0.60) = P(Z \leq \frac{0.60-0.70}{0.1}) = P(Z \leq -1) = 0.1587 > 0.10$. Condition has not been met. We do not reject the null at $\alpha = 0.10$ significance level. Hypothesis test is inconclusive. Optional additional conclusion: what we got is probably not significantly different from what we ordered.

Old Questions Relevant for Regression, ANOVA, Chi Square

Question 1 - Previous Class Exercise

1) You are given these summary statistics: $r = 0.60$, $s_y = 10$, $s_x = 2$, $\bar{X} = 10$, $\bar{Y} = 5$. Write down the regression line.

2) Predict the value of Y when $X = 10$.

3) Predict the increase in Y when X is increased by 5.

Answers

1) Slope coefficient $b = r\frac{s_y}{s_x} = 0.60 \cdot (10/2) = 3$. Intercept $a = 5 - 3 \cdot 10 = -25$. So, $\hat{Y} = 3X - 25$.
2) $\hat{Y} = 30 - 25 = 5$.
3) $b =$ increase in Y for every 1 increase in X. So, answer is $5b = 15$.

Question 2 - Previous Class Exercise

Suppose the sampling distribution is normally distributed with standard error 1. The one-tailed test null hypothesis is that the mean of the sampling disitrubtion is $\mu = 0$, and the alternative hypothesis is $\mu > 0$. Significance level is $\alpha = 0.10$ and $P(Z \geq 1.28) = 0.10$. Suppose that the true mean is $\mu_{true} = 2.28$.

Recall that power is the probability of rejecting a false null hypothesis. What is the power of this hypothesis test?

Answers

Draw the normal distribution with mean $\mu_{true} = 2.28$. Mark 1.28 and see that it is 1 standard error to the left of this true mean. The area $P(Z \geq -1) = 1 - 0.1587 = 0.8413$ is the power.

Question 3 - Previous Class Exercise

You are given the fitted regression line: $\hat{Y} = 3X + 5$. And that your estimate of the slope coefficient has standard error $s_b = 0.4$ There are $n = 10$ pairs of data in your bivariate data set. Perform a hypothesis test on $H_0 = 4$ and $H_A < 4$, at $\alpha = 0.03$ level of significance. Write a conclusion.

Answers

We know from class this is a t-test with degrees of freedom $n - 2 = 8$. The t-statistic is $\frac{3-4}{0.4} = -2.5$. So, we look at the t-tables for the area under the curve to get $P($ sample slope $\leq -2.5) < P($ sample slope $\leq 2.31) = 0.025 < \alpha$. Condition has been met, so the null is rejected at 0.03 level of significance. The true slope coefficient is probably less than 4.

Question 4 - Previous Class Exercise

You have $k = 3$ samples of size $n = 6$ each. Each of these samples are drawn from a normal distribution with the same variances but possibly different means μ_1, μ_2, μ_3 respectively. Sample variances : $s_1^2 = s_2^2 = s_3^2 = 10$. (entirely by chance I assure you :D)

The (sample) variance of the 3 sample means is 7. Recall the formulas: MSB $= n \times$ variance of the means, MSE $= \frac{1}{k}(s_1^2 + s_2^2 + s_3^2)$. And $df_1 = df_{numerator} = (k-1)$, $df_2 = df_{denominator} = (N-k)$, where N is the total number of data points in all k samples.

Write down the null and alternative hypothesis, then perform an ANOVA hypothesis test at $\alpha = 0.05$. Write a conclusion.

Answers

Null: $\mu_1 = \mu_2 = \mu_3$, alternative: some means are different. $F = MSB/MSE = (6 \cdot 7)/(30/3) = 42/10 = 4.2$. F table with $df_1 = 2$ and $df_2 = 15$ says that $P(F \geq 4.2) < P(F \geq 3.68) = 0.05 = \alpha$. Alternatively, you could say that the F-statistic is 4.2, which is greater than the critical F-value of 3.68. Condition has been met, so the null is rejected at 0.05 level of significance. Some means are probably different.

Old Questions Relevant for Sampling Distributions, Confidence Intervals, Hypothesis Testing

Previous questions that are relevant for hypothesis testing are found in **Homework 4** and **Homework 5**. Make sure you can do them before you take the final exam!

There are also a lot of relevant questions from the midterm exam and midterm practice questions. These are reproduced here with their answers for your convenience.

Midterm Question 4 (10 points)

A population has mean age 30, with variance 64. The distribution of ages is highly negatively skewed, with a long tail to the left.

a) Are the mean age and the standard deviation good summary statistics for the distribution of ages in this population? Explain. *(2 points)*

No, 1 pt (reasons must be right). The distribution is highly negatively skewed 1 pt. Alternative: This is not symmetric. Mean and standard deviation would be good if it was symmetric. Will take away 1 pt for saying anything that is false but still got the general idea.

b) Which distribution would be a good approximation for the sampling distribution of the means in samples of size 16? Also state the mean and standard error of this approximation. Simplify your answer as much as possible. *(3 points)*

Normal distribution (1 pt) with mean 30 (1 pt) and standard error 2 (1 pt).

c) What theorem allowed us to make the approximation in the previous part b)? *(1 point)*

Central Limit Theorem. Get the 1 pt for stating this, even if did not apply it to part b), or applied it incorrectly.

d) If someone took a sample of size 16 from this population, and calculated the mean age in the sample, what is the approximate probability that the mean age would be in the interval $[28.6, 34.4]$? Show your work. You may do your calculations with 3 significant figures. *(4 points)*

Hint: illustrations are allowed. It might be helpful to draw the curve and shade the relevant area underneath.

Using the z-table: $P(z\text{-score} \leq 2.2) = 0.986$ 1 pt, $P(z\text{-score} \leq -0.7) = 0.242$, 1 pt. Awarded for each correct z-score calculation.

Then, $0.986 - 0.242 = 0.744$, 2 pts. If you managed to explain or show that you understand the logic of this last step, you get 2 pts even if the z-score calculations above are not correct, due to calculation errors or careless mistakes etc.

Midterm Question 5 (10 points)

a) A truck is full of watermelons that has population mean weight of 9 kilograms (kg). You do not know the variance. However, you know that the weights of the watermelons are normally distributed.

Suppose you take a simple random sample of $n = 4$ watermelons, and calculated a sample mean of $M = 8$ kg from this sample. Using this sample, you got an estimate of the standard deviation $s = 2$ kg.

Construct a 90 % confidence interval for the mean weight of watermelons.

Show your work. Simplify your answer as much as possible so that the interval is of the form $[a, b]$, where a and b are numbers to 2 signficant figures *(6 points)*.

Degrees of freedom $= 4 - 1 = 3$, 1 pt.

t-value from the table for 90 % confidence interval and degrees of freedom 3 is $t = 2.35$, 2 pts.

Standard error, $\frac{s}{\sqrt{n}} = \frac{2}{\sqrt{4}} = \frac{2}{2} = 1$, 1 pts.

90 % confidence interval, $[8 - 2.35 \cdot \frac{2}{\sqrt{4}}, 8 + 2.35 \cdot \frac{2}{\sqrt{4}}] = [5.65, 10.35]$, 2 pts.

You get points for calculating the correct standard error, even if the entire set up is wrong. E.g. using the normal distribution here. BUT only if the standard error is correct.

However, you need to be using the correct formula and set up to get the 2 pts for the confidence interval. You will get these 2 pts even if you made calculation errors previously.

b) Consider a population of males and females. You do not know the proportion of females in the population.

Suppose you take a simple random sample of $n = 4$ people from this population, and found that the proportion of females in this sample is $p = \frac{1}{5}$.

Construct the 95 % confidence interval for the population proportion of females.

Show your work. Simplify your answer as much as possible, so that the interval is of the form $[a, b]$, where a and b are might contain a mix of numbers, and fractions that are reduced to lowest terms. You do not have to add/subtract $\frac{0.5}{n}$ from the upper and lower limit of the interval. *(4 points)*

Normal approximation to the binomial gives the z-value as $z = 1.96$. You can ue the 68-95 heuristic to get $z = 2$, which is fine. 1 pt.

Standard error, $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{\frac{1}{5}\frac{4}{5}}{4}} = \sqrt{\frac{1}{25}} = \frac{1}{5}$, 2 pts.

$[\frac{1}{5} - 1.96 \cdot \frac{1}{5}, \frac{1}{5} + 1.96 \cdot \frac{1}{5}]$, 1 pt

Same as before, you get points for correct standard error, even if the entire set up is wrong. E.g. using the normal distribution here. BUT only if the standard error is correct.

Likewise, you need to be using the correct formula and set up to get the 1 pt for the confidence interval. You will get these 1 pt even if you made calculation errors previously.

Practice Midterm Question 5

The weights for 1000 bars of gold are well approximated by a normal distribution, with mean 100 grams and standard deviation 20 grams. You may round your final answers to the nearest integer.

a) Approximately how many bars of gold weigh between 100 grams and 130 grams inclusive? (2 pts)

b) Approximately how many bars of gold weigh less than 80 grams?

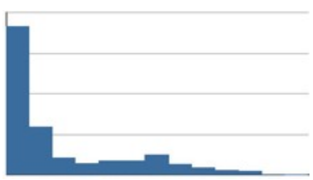
Answer: Let X be the weight of a randomly selected bar of gold. $P(X \leq k) = P(Z \leq \frac{k-100}{20})$, where Z is standard normal (z-tables).

a) So, $P(X \leq 100) = P(Z \leq \frac{100-100}{20} = 0) = 0.50$, and $P(X \leq 130) = P(Z \leq \frac{130-100}{20}) = P(Z \leq 1.5) = 0.9332$. So, $P(X \leq 130) - P(X \leq 100) = 0.9332 - 0.50 = 0.4332$. Ans: approximately $0.4332 \cdot 1000 \approx 433$ bars.

b) Also, $P(X \leq 80) = P(Z \leq \frac{80-100}{20}) = P(Z \leq -1) = 0.1587$. Ans: $0.1587 \cdot 1000 \approx 159$.

Practice Midterm Question 6

The incomes of people in country X has a distribution that looks like the one below, with population mean μ and variance σ^2.



a) If you take a large simple random sample of n incomes from country X, what is a good approximation of the sampling distribution of the sample mean M? What are the mean and variance of this approximation? (2 pt)

b) Which theorem made the approximation in the previous question possible? (1 pt)

c) If another researcher independently took another large simple random sample of n incomes from country X, what is the probability that his sample mean would be in the interval $[\mu - \frac{\sigma}{\sqrt{n}}, \mu + \frac{\sigma}{\sqrt{n}}]$? (2 pts)

Answer: a) Normal distribution with mean μ and variance $\frac{\sigma^2}{n}$. b) Central Limit Theorem. c) Using the 68-95 heuristic/rule from the textbook, the probability is 0.68.

Practice Midterm Question 7

You are a small employee in a big chain of supermarkets. Your boss wants to know if the companys image is attracting more men than women, or is it roughly 50/50. Population has two types: Men and Women. You take a sample of size $n = 10$. You find that your sample proportion of men is 60 % or 0.60 or $\frac{3}{5}$.

You want to construct a numerical 90 % confidence interval for the population proportion of men, because you took Math 10 before. How would you do it?

Your answer must be in the form $[a, b]$, where a and b are numbers that may include square roots. You do not have to evaluate square roots. You do not have to adjust the limits of the interval by $\frac{0.5}{n}$. *(6 pts).*

Answer: formula for the confidence interval is $[p - Z_\alpha S_p, p + Z_\alpha S_p]$, where $p = \frac{3}{5}$, $Z_\alpha = 1.64$, and the standard error $S_p = \sqrt{\frac{\frac{3}{5} \cdot \frac{2}{5}}{10}} = \sqrt{\frac{6}{250}}$.

Practice Midterm Question 8

Suppose you have a block of metal that is exactly 500 grams in weight. Sorry, no imperial units allowed in my class.

You have an electronic weighing scale that may or may not be faulty. You put this block of metal on the weighing scale $n = 25$ times, and record the results.

Each result varies a little due to various reasons (positioning, random mechnical errors etc), but you are hoping to find out if the scale is correct on average, or systematically giving you lower/higher than the actual weight.

If the scale is correct on average, it would produce results drawn from a normal distribution with mean 500 grams and unknown variance σ^2. (E.g. real life scales will tell you they are intended to be accurate within X grams)

The sample mean of your $n = 25$ data points is $M = 490$ grams.

1. Can you conclude that the electronic weighing scale is faulty and is systematically giving you results that are lower than the actual weight, on average? Since $M = 490 < 500$ true weight? Explain. *(2 pts)*

2. Using this set of sample data, you calculated an estimate of the standard deviation $s = 20$ grams. What sampling distribution of the mean would you use? State all the parameters of this sampling distribution (using μ for the real mean). Why do you use this sampling distribution? *(3 pts)*

3. Using the statistics produced by your sample, construct a 95 % confidence interval for the mean. *(4 pts)*

Answers

1) No. Even if the scale is correct on average, you could have gotten 490 by chance.

2) No population variance, have to use t distribution. Mean μ, standard error $\frac{20}{\sqrt{25}} = 4$, degrees of freedom 24.

3) Degrees of freedom $= 25 - 1 = 24$. Find the t-value, $t = 2.06$. Formula: $[M - t \cdot SE, M + t \cdot SE]$, where SE is the standard error given.

Relevant Old Class Exercise

X is a normally distributed variable with (population) mean 5 and variance 36.

1) Using the z-tables, calculate the probability that X will be in the interval $[-4.6, 9.8]$.
Hint: find $P(-4.6 \leq X \leq 9.8)$.

2) If I want an interval symmetric around the mean, $[5-a, 5+a]$, so that the probability that X is in this interval is 0.8444, what would a be? Hint: find a so that $P(5-a \leq X \leq 5+a) = 0.8444$. The answer must be numerical, and can contain products, but you do not have to simplify your answer.

3) A researcher generates a sample of size $n = 36$ consisting of independent values $X_1, X_2, ..., X_{36}$ from the normal distribution above. She calculates the mean of this sample and write it as $\bar{X}$.

In general, what is the distribution of the sample means $\bar{X}$ generated like this? Or rather (same question), what is the sampling distribution of the mean? What is the mean and variance of this sampling distribution? Hint: technically you don't need CLT here, but using CLT is fine too.

4) The standard deviation of the distribution you found in part 3) is call the standard error. Using the z-tables, calculate the probability that $\bar{X}$ will be in the interval $[3.9, 6.1]$.

5) Suppose you too take a new sample of size $n = 36$, like what the researcher did in part 3). In your case, you got $\bar{X} = 7$. Construct a 90% confidence interval for the mean.

6) Look at the interval you got in part 5). It is nowhere near the mean of 5. Why is it still a "90" percent confidence interval, when the mean is clearly not in it?

Answers

1) $P(-4.6 \leq X \leq 9.8) = P(X \leq 9.8) - P(X \leq -4.6) = P(Z \leq 0.8) - P(Z \leq -1.8) = 0.7881 - 0.0548 = 0.7333$.

2) $a = 1.42 \cdot 6$, we got 1.42 from the z-tables.

3) It is fine to say by CLT it is normal or approximately. The mean is 5. The variance is $\frac{\sigma^2}{n} = \frac{36}{36} = 1$.

4) 0.7286.

5) Use the formula $[\bar{X} - +z \cdot 1] = [7 - 1.64, 7 + 1.64] = [5.36, 8.64]$.

6) The entire procedure has a 0.90 chance of producing an interval that contains the mean. Taking a sample of size $n = 36$ and calculating a new $\bar{X}$ is part of the procedure.