Here are the answers to the Data Science Master's assignment:

**Q1: Difference between a t-test and a z-test**

* **t-test**: Used when the sample size is small (n < 30) and the population standard deviation is unknown. It is based on the t-distribution, which accounts for the added variability in small samples.
* **z-test**: Used when the sample size is large (n ≥ 30) or the population standard deviation is known. It is based on the normal distribution.

**Example scenario**:

* **t-test**: A study tests the average height of a sample of 15 individuals from a population where the standard deviation of height is unknown.
* **z-test**: A quality control test of 1000 manufactured items, where the population standard deviation is known.

**Q2: One-tailed vs Two-tailed tests**

* **One-tailed test**: A hypothesis test where the critical area is only in one tail of the distribution (e.g., testing if a sample mean is greater than a hypothesized value).
* **Two-tailed test**: A hypothesis test where the critical area is split between both tails of the distribution (e.g., testing if a sample mean is different from a hypothesized value).

**Example**:

* **One-tailed test**: Testing if a new drug decreases blood pressure by more than 10 mmHg.
* **Two-tailed test**: Testing if a new drug changes blood pressure by any amount (increase or decrease).

**Q3: Type 1 and Type 2 errors in hypothesis testing**

* **Type 1 error** (False Positive): Rejecting the null hypothesis when it is actually true.
  + **Example**: Concluding a new drug is effective when it is actually not.
* **Type 2 error** (False Negative): Failing to reject the null hypothesis when it is actually false.
  + **Example**: Concluding a new drug is ineffective when it is actually effective.

**Q4: Bayes' Theorem**

Bayes' Theorem provides a way to update the probability of a hypothesis based on new evidence.

The formula is:

P(A∣B)=P(B∣A)⋅P(A)P(B)P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}

Where:

* P(A∣B)P(A|B) is the probability of AA given BB.
* P(B∣A)P(B|A) is the probability of BB given AA.
* P(A)P(A) and P(B)P(B) are the prior probabilities of AA and BB.

**Example**: If 1% of people have a disease, and a test has a 95% sensitivity (true positive rate) and 90% specificity (true negative rate), Bayes' Theorem helps to calculate the probability of having the disease given a positive test result.

**Q5: Confidence Interval**

A confidence interval (CI) is a range of values used to estimate the true population parameter. It is calculated as:

CI=μ^±Z×σnCI = \hat{\mu} \pm Z \times \frac{\sigma}{\sqrt{n}}

Where:

* μ^\hat{\mu} is the sample mean.
* ZZ is the z-score corresponding to the desired confidence level.
* σ\sigma is the population standard deviation.
* nn is the sample size.

**Example**: For a sample of 100 students with an average score of 75 and a standard deviation of 10, a 95% confidence interval for the mean would be calculated using the formula above.

**Q6: Bayes' Theorem for an Event**

Given prior probability P(A)=0.3P(A) = 0.3, likelihood P(B∣A)=0.8P(B|A) = 0.8, and evidence probability P(B)=0.5P(B) = 0.5, Bayes' Theorem helps to calculate the posterior probability P(A∣B)P(A|B).

Using the formula:

P(A∣B)=P(B∣A)⋅P(A)P(B)=0.8×0.30.5=0.48P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{0.8 \times 0.3}{0.5} = 0.48

**Q7: 95% Confidence Interval**

Given:

* Sample mean μ^=50\hat{\mu} = 50
* Standard deviation σ=5\sigma = 5
* Sample size n=30n = 30

Using the formula for the confidence interval:

CI=μ^±Z×σnCI = \hat{\mu} \pm Z \times \frac{\sigma}{\sqrt{n}}

For a 95% confidence level, Z=1.96Z = 1.96:

CI=50±1.96×530=50±1.79CI = 50 \pm 1.96 \times \frac{5}{\sqrt{30}} = 50 \pm 1.79

So, the 95% confidence interval is approximately (48.21, 51.79).

**Q8: Margin of Error and Sample Size**

The margin of error (ME) is the maximum expected difference between the sample estimate and the true population parameter:

ME=Z×σnME = Z \times \frac{\sigma}{\sqrt{n}}

A larger sample size decreases the margin of error, as the standard error (σn\frac{\sigma}{\sqrt{n}}) decreases with increasing nn.

**Example**: For a sample of 50 versus 200, the margin of error for the larger sample will be smaller.

**Q9: Z-Score Calculation**

Given:

* Value X=75X = 75
* Population mean μ=70\mu = 70
* Population standard deviation σ=5\sigma = 5

The formula for the z-score is:

Z=X−μσ=75−705=1Z = \frac{X - \mu}{\sigma} = \frac{75 - 70}{5} = 1

Interpretation: The value 75 is 1 standard deviation above the population mean.

**Q10: Hypothesis Test for Weight Loss Drug**

Given:

* Sample size n=50n = 50
* Sample mean Xˉ=6\bar{X} = 6
* Standard deviation s=2.5s = 2.5
* Significance level α=0.05\alpha = 0.05

We perform a one-sample t-test with the null hypothesis H0:μ=0H\_0: \mu = 0 (no effect) and alternative H1:μ>0H\_1: \mu > 0.

Calculate the t-statistic:

t=Xˉ−μ0s/n=6−02.5/50≈17.89t = \frac{\bar{X} - \mu\_0}{s / \sqrt{n}} = \frac{6 - 0}{2.5 / \sqrt{50}} \approx 17.89

With tt greater than the critical value (from the t-distribution table), we reject the null hypothesis and conclude the drug is effective.

**Q11: Confidence Interval for Proportion**

Given:

* Sample size n=500n = 500
* Proportion p=0.65p = 0.65

The 95% confidence interval for a population proportion is calculated using the formula:

CI=p±Z×p(1−p)nCI = p \pm Z \times \sqrt{\frac{p(1-p)}{n}}

For p=0.65p = 0.65, Z=1.96Z = 1.96, and n=500n = 500, we get:

CI=0.65±1.96×0.65(0.35)500≈0.65±0.055CI = 0.65 \pm 1.96 \times \sqrt{\frac{0.65(0.35)}{500}} \approx 0.65 \pm 0.055

So, the 95% confidence interval is approximately (0.595, 0.705).

**Q12: Hypothesis Test for Teaching Methods**

Given:

* Sample A: mean 8585, SD 66, n=30n = 30
* Sample B: mean 8282, SD 55, n=30n = 30
* Significance level α=0.01\alpha = 0.01

We conduct a two-sample t-test. The t-statistic is calculated, and based on the result, we decide whether to reject the null hypothesis that the means are equal.

**Q13: 90% Confidence Interval for Population Mean**

Given:

* Sample mean Xˉ=65\bar{X} = 65
* Population mean μ=60\mu = 60
* Standard deviation σ=8\sigma = 8
* Sample size n=50n = 50

Using the formula for the confidence interval:

CI=Xˉ±Z×σn=65±1.645×850≈(63.09,66.91)CI = \bar{X} \pm Z \times \frac{\sigma}{\sqrt{n}} = 65 \pm 1.645 \times \frac{8}{\sqrt{50}} \approx (63.09, 66.91)

**Q14: Hypothesis Test for Caffeine and Reaction Time**

Given:

* Sample mean Xˉ=0.25\bar{X} = 0.25
* Standard deviation s=0.05s = 0.05
* Sample size n=30n = 30
* Significance level α=0.10\alpha = 0.10

We perform a one-sample t-test to determine if caffeine has a significant effect on reaction time. The t-statistic is calculated and compared to the critical value to decide whether to reject the null hypothesis.