CQE Academy

CQE EXAM

Statistics Practice Exam

25 Questions That Will Challenge Your Knowledge of Stats!

ANDY ROBERTSON

Statistics!

The scariest (and most difficult) topic on the CQE Exam is **Statistics**.

The CQE Body of Knowledge have 8 major Statistics (Quantitative Methods and Tools) chapter:

- **Collecting & Summarizing Data**
- **Probability (Quantitative Methods)**
- **Probability Distribution**
- **Statistical Decision Making**
- **Relationships Between Variables**
- **Statistical Process Control**
- **Process & Performance Capability**
- **Design and Analysis of Experiments**

Free Practice Exam

To help you assess how well you know Statistics, I put together this free 25 question practice exam.

I've also added the solutions to these questions starting down on page 11.

Feedback!

Please enjoy the quiz, and send me any feedback at **Andy@CQEAcademy.com**.

Practice Exam for Statistics from CQEAcademy.com

1. **Identify the correct equation below associated with the sample variance calculation:**

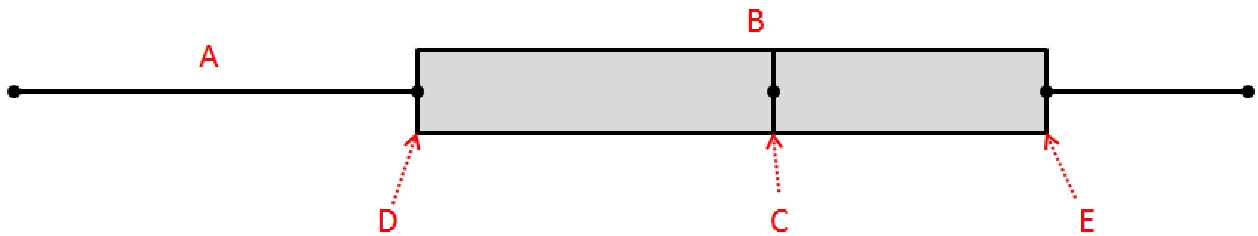
- $s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$
- $\sigma^2 = \frac{\sum(x-\bar{\mu})^2}{N}$
- $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$
- $\sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1}$
- $s^2 = \frac{\sum(x-\bar{\mu})^2}{N}$
- $s^2 = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$

2. **Calculate the sample standard deviation of the following data set: 2, 4, 6, 8**
   - 6.66
   - 5
   - 2.24
   - 2.58

3. **If you flip 3 coins simultaneously, what is the probability that you only get 1 coin to land on heads:**
   - 12.5%
   - 25.0%
   - 37.5%
   - 50.0%
   - 62.5%

4. Match the following Terms with their Location on the Box Plot Below (A – E):

- Median Point
- Whisker
- Box
- Upper Quartile
- Q_1



5. A shipping operation distributed product at a mean time of 48 hours from receipt of order with a standard deviation of 6 hours. What percentage of shipments go out between 42 - 54 hours from time of receipt:

- 34%
- 68%
- 66%
- 32%

6. What is the critical z-value associated with a 2-sided confidence interval that's associated with a 20% alpha risk?

- z-score = 1.29
- z-score = 1.65
- z-score = 1.72
- z-score = 1.34

7. **You've sampled 60 units from the latest production lot to measure the width of the product. The sample mean is 6.75in and the population standard deviation is known to be 0.75in.**

   Calculate the 95% confidence interval for the population mean:

   - 6.75 + 0.219
   - 6.75 + 1.470
   - 6.75 + 0.024
   - 6.75 + 0.189

8. **You've measured 8 units from the latest production lot to measure the length of the parts. You calculate the sample mean to be 16.5in, and the sample standard deviation to be 1.5 in.**

   Calculate the 80% confidence interval for the population standard deviation.

   - $1.224 < \sigma < 2.521$
   - $1.145 < \sigma < 2.358$
   - $1.086 < \sigma < 2.124$
   - $1.310 < \sigma < 5.559$

9. **You manufacture a widget whose average length is 4.20 inches. You've upgraded your manufacturing equipment and you believe that it will not impact the overall length of the part.**

   You know the population standard deviation is 0.10 inches, and the sample mean of the 40 parts you measured is 4.24 inches. Using a 5% significance level to determine if the average length of the part has changed. Assume the length of the part is normally distributed.

   Identify all of the statements below that are true:

   - The null hypothesis, H_0: $\mu = 4.24$ inches
   - The alternative hypothesis, H_a: $\mu \neq 4.20$ inches
   - The hypothesis test is a 1-sided test
   - The critical rejection is $t_{crit} = 1.96$
   - The test statistic is $z_{stat} = 2.53$
   - The result of the test is the failure to reject the null hypothesis

10. **The One-Way ANOVA Analysis below has 10 treatment groups with the total degrees of freedom of 19.**

Variation Source	Sum of Squares (SS)	Degrees of freedom (DF)	Mean Squares (MS)	F-Value
Treatment (Between)				
Error (Within)	55			
Total	100	19		

Calculate the Treatment Mean Square for this ANOVA Table.

- 4.5
- 5
- 5.5
- 6.1

11. **You're creating a linear regression model for your data and you've calculated the following values.**

$$S_{yy} = 31, S_{xy} = 12, S_{xx} = 76$$

Based on these results, what percentage of variation in Y, can be explained by the variation in X.

- 100%
- 81%
- 76%
- 43%
- 28%
- 13%
- 6%
- 1%
- Not Enough Information Provided

12. **You're creating a linear regression model for your data and you've calculated the following values.**
S_{yy} = 4130, S_{xy} = 1527, S_{xx} = 626.86, β_0 = 17.81

What is the predicted value of Y when X = 23.

- 48
- 54
- 56
- 66
- 74
- Not Enough Information Provided

13. **You're creating a linear regression model for your data and you've calculated the following values.**
S_{yy} = 31, S_{xy} = 12, S_{xx} = 76

What is correlation coefficient for this data set:

- -1
- -0.4
- 0
- 0.25
- 0.72
- 0.95
- 1
- Not Enough Information Provided

14. **Which probability distribution is used to construct the c-chart:**
- The Normal Distribution
- The Exponential Distribution
- The Poisson Distribution
- The Binomial Distribution

15. What type of control chart would be used to monitor the number of defectives for a process with a constant sample size:

- P Chart
- NP Chart
- C Chart
- U Chart

16. The Fake News Media Corporate collected data on the number of fake news stories published every day and constructed a p-chart.

A sample of 100 articles are inspected every day, however this can vary. The average percentage of fake news stories (defectives) was calculated as 0.106.

On a particular day, 200 articles were inspected and 47 fake news reports were observed. What is the conclusion of this day:

- The sample is in statistical control and this is a normal level of fake news
- The sample is out of statistical control and there's a lot of fake news going around

17. A p-chart monitors what type of attribute:

- The number of defective items in a subgroup
- The number of defects in a subgroup
- The percentage of defects in a subgroup
- The percentage of defectives in a subgroup

18. What is the UCL for a p-chart when the average daily inspection quantity is 50, and the historical percentage of defectives is 0.05?

- 0.21
- 0.09
- 0.29
- 0.14
- 0.17

19. **You're constructing an NP chart, where you've sampled from 25 subgroups, each with 100 samples, and found a total of 145 defective units.**

    Calculate the UCL for this process.
    - 5.8
    - 0.058
    - 7.0
    - 12.8
    - 14.5
    - Not Enough Information Provided

20. **You're manufacturing a widget and using an X-bar and R chart to control the critical feature of the product. Your normal process has the following attributes:**

    X-double bar is 225, R-bar is 12, n = 8.

    Identify the upper and lower control limits for the X-bar chart:

    - 0
    - 220.52
    - 229.48
    - 1.63
    - 233.14
    - 218.71
    - 22.37

21. **Calculate Cpk for the following Parameters: (USL = 1.35, LSL = 1.15, σ = 0.025, μ = 1.25)**
    - 0.67
    - 1.0
    - 1.33
    - 1.67
    - 2.0

22. Calculate Cpk for the following Parameters: (USL = 205, LSL = 145, σ = 10, μ = 190)

- 0.50
- 0.67
- 1.0
- 1.33
- 1.50

23. What Cpk value will theoretically result in 1 defect per million?

- 1.0
- 1.33
- 1.66
- 2.0
- 6.0

24. How many treatments would be required for a DOE with 8 factors where a quarter factorial design is chosen:

- 256
- 128
- 64
- 32
- 16
- 8

25. You performed a full factorial DOE to improve the yield of a process with two factors at two levels and have measured the following response values.

What is the estimated effect of Factor B:

		Factors		Response
		A	B	% Yield
Treatments	1	+	+	64
	2	-	+	75
	3	+	-	87
	4	-	-	95

- -9.5
- -21.5
- 11
- -1.5
- -8

Solutions for Practice Exam of Statistics

1. **Identify the correct equation below associated with the sample variance calculation:**

- $s^2 = \frac{\sum(x-\overline{x})^2}{n-1}$
- $\sigma^2 = \frac{\sum(x-\overline{\mu})^2}{N}$
- $s = \sqrt{\frac{\sum(x-\overline{x})^2}{n-1}}$
- $\sigma^2 = \frac{\sum(x-\overline{x})^2}{n-1}$
- $s^2 = \frac{\sum(x-\overline{\mu})^2}{N}$
- $s^2 = \sqrt{\frac{\sum(x-\overline{x})^2}{n-1}}$

This question is all about the **sample variance** and it is an attempt to highlight the *differences between the sample variance & the population variance*; along with the differences between variance & standard deviation.

The first thing to remember is that the sample variance is always described as s^2 while σ^2 always applies to the population variance - so we can immediately exclude those equations shown in **red**.

Then, remember that the standard deviation equation is the square root of variance. So, any equation that includes the square root would not be used to calculate variance. So, we can immediately exclude the bottom equation in **purple**.

Lastly, what you must know is that when calculating the sample variance, you use the sample mean, not the population mean, and you also divide by n-1, not N, so we can exclude the equation in **orange**. This leaves us with the sample variance equation shown in **green**.

2. **Calculate the sample standard deviation of the following data set: 2, 4, 6, 8**
   - 6.66
   - 5
   - 2.24
   - **2.58**

Notice here that we're talking about a sample not a population, *which has an impact on the way variance is calculated.*

$$\textbf{\textit{Sample Standard Deviation}} = s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

$$\textbf{\textit{sample mean}} = \bar{x} = \frac{\sum x}{n} = \frac{2+4+6+8}{4} = \frac{20}{4} = 5$$

So now we can calculate variance & standard deviation - let's do this using a table:

(x)	$(x-\bar{x})$	$(x-\bar{x})^2$
2	(2 - 5) = -3	9
4	(4 - 5) = -1	1
6	(6 - 5) = 1	1
8	(8 - 5) = 3	9
		$\sum(x-\bar{x})^2 = 20$

Notice in this table the left-hand column is the individual measurement values (X).

Then in the middle column we calculate the deviation of each observation from the average value.

This deviation from the mean is then squared in the 3rd column so that we can ultimately sum up that squared deviation.

$$\textbf{\textit{Sample Standard Deviation}} = s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum(9+1+1+9)}{4-1}} = \sqrt{\frac{20}{3}} = 2.58$$

3. **If you flip 3 coins simultaneously, what is the probability that you only get 1-coin land on heads:**
   - **12.5%**
   - **25.0%**
   - **37.5%**
   - **50.0%**
   - **62.5%**

Let's start this problem solving by determining our **experiments sample space** - this is the total combination of all possible outcomes.

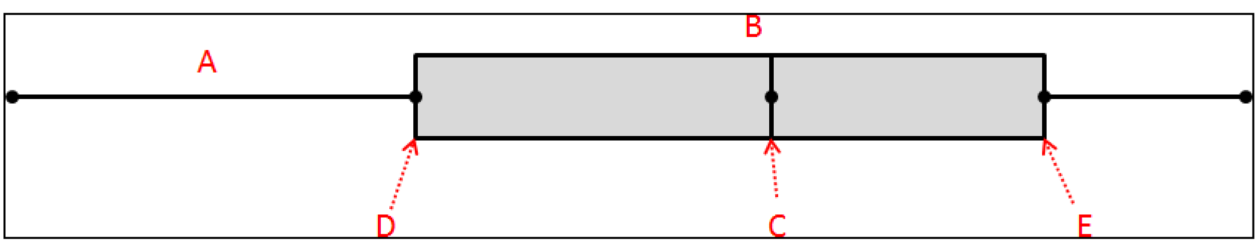
Which for this experiment looks like this: (HHH, HHT, HTH, HTT, THH, THT, TTH, TTT).

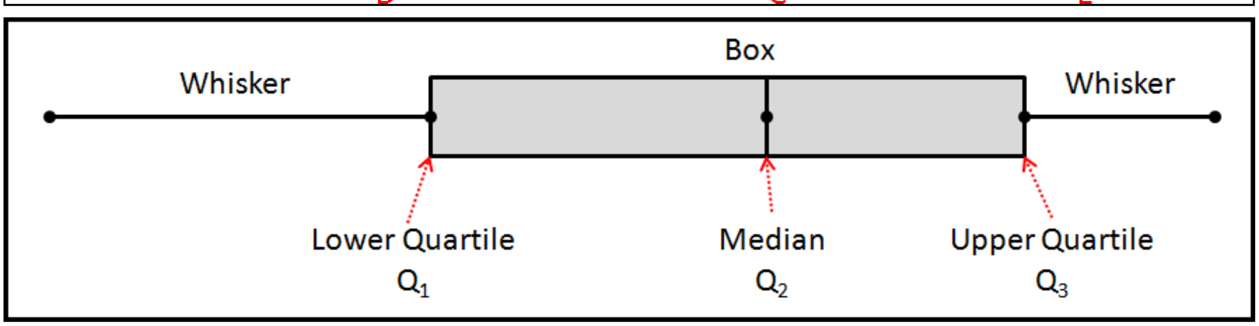
So, there are 8 possible outcomes.

Now we can define our Event (A) as all outcomes where only 1 coin lands on heads, which include (HTT, THT, TTH), so 3 out of the 8 outcomes within the sample space are part of our event.

$$\textit{The Probability of Event A} = \frac{\textit{The \# of Outcomes in A}}{\textit{The Total \# of Possible Outcomes}} = \frac{3}{8} = 37.5\%$$

4. **Match the following Terms with their Location on the Box Plot Below:**
   - A – Whisker
   - B – Box
   - C – Median Point
   - D – Q_1
   - E – Upper Quartile



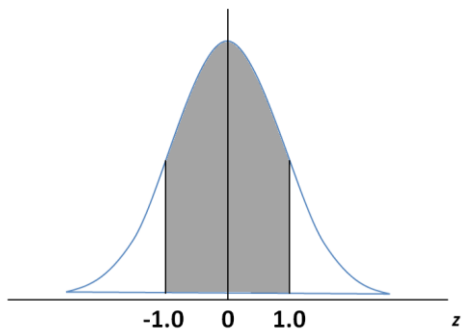


5. **A shipping operation distributed product at a mean time of 48 hours from receipt of order with a standard deviation of 6 hours. What percentage of shipments go out between 42 - 54 hours from time of receipt. Reference the Probability Tables from NIST.**
   - 34%
   - 68%
   - 66%
   - 32%

The first thing we must do is to calculate the Z transformation for the two time values (42 hours & 54 hours).

$$Z - Transformation: Z(x) = \frac{X - \mu}{\sigma}$$

$$\boldsymbol{Z(X = 42\ hours) = \frac{42 - 48}{6} = \frac{-6}{6} = -1.0} \quad AND \quad \boldsymbol{Z(X = 54\ Hours) = \frac{54 - 48}{6} = \frac{6}{6} = 1.0}$$



The probability that Z is between Z = -1.0 & 1.0 is equal to the shaded area above. We can lookup the area under this curve using the Probability Tables from NIST where we can find that the probability of Z = 1.0 is equal to 0.34134, which is also equal to the probability of Z = -1.0.

Area under the Normal Curve from 0 to X

X	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41308	0.41466

So, the total probability from z= -1 to z = 1 is **68.168%**, which represents the shaded area under the normal curve above.

6. **What is the critical z-value associated with a 2-sided confidence interval that's associated with a 20% alpha risk.**

   [NIST Z-Table for Normal Distribution](#)

- **z-score = 1.29**
- ~~z-score = 1.65~~
- ~~z-score = 1.72~~
- ~~z-score = 1.34~~

Because it's a 2-sided distribution, and our alpha risk is 20%, we must look for the z-score that's associated with the area under the curve of 0.400, which is equal to 40% of the distribution.

This would capture 40% on the left half & right half of the distribution, leaving the remaining 20% of the alpha risk in the rejection area of the tails of the distribution.

The closest z-score associated with 0.400 probability is z = 1.29

Area under the Normal Curve from 0 to X

X	0.00	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00000	0.01595	0.01994	0.02392	0.02790	0.03188	0.0[illegible]586
0.1	0.03983	0.05567	0.05962	0.06356	0.06749	0.07142	0.0[illegible]535
0.2	0.07926	0.09483	0.09871	0.10257	0.10642	0.11026	0.1[illegible]409
0.3	0.11791	0.13307	0.13683	0.14058	0.14431	0.14803	0.1[illegible]173
0.4	0.15542	0.17003	0.17364	0.17724	0.18082	0.18439	0.1[illegible]793
0.5	0.19146	0.20540	0.20884	0.21226	0.21566	0.21904	0.2[illegible]240
0.6	0.22575	0.23891	0.24215	0.24537	0.24857	0.25175	0.2[illegible]490
0.7	0.25804	0.27035	0.27337	0.27637	0.27935	0.28230	0.2[illegible]524
0.8	0.28814	0.29955	0.30234	0.30511	0.30785	0.31057	0.3[illegible]327
0.9	0.31594	0.32639	0.32894	0.33147	0.33398	0.33646	0.3[illegible]891
1.0	0.34134	0.35083	0.35314	0.35543	0.35769	0.35993	0.3[illegible]214
1.1	0.36433	0.37286	0.37493	0.37698	0.37900	0.38100	0.3[illegible]298
1.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.40147
1.3	0.40320	0.40988	0.41149	0.41308	0.41466	0.41621	0.41774
1.4	0.41924	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189

7. **You've sampled 60 units from the latest production lot to measure the width of the product. The sample mean is 6.75in and the population standard deviation is known to be 0.75in. Calculate the 95% confidence interval for the population mean:**

- ~~6.75 ± 0.219~~
- ~~6.75 ± 1.470~~
- ~~6.75 ± 0.024~~
- 6.75 ± 0.189

Ok, we know after reading the question:

- **n = 60,**
- **σ = 0.75in,**
- **α = 0.05,**
- **x-bar = 6.75in**

Because we've sampled more than 30 units and the population standard deviation is known, we can use the Z-score approach to this confidence interval problem.

We need to find the Z-score associated with the 95% confidence interval using the NIST Z-Table, where we find Z = 1.96.

$$\textit{Interval Estimate of Population Mean (known variance)} : \bar{x} \pm Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

$$\textit{Interval Estimate} : 6.75 \pm 1.96 * \frac{0.75}{\sqrt{60}}$$

$$\textit{Interval Estimate} : 6.75 \pm 0.189$$

8. **You've measured 8 units from the latest production lot to measure the length of the parts. You calculate the sample mean to be 16.5in, and the sample standard deviation to be 1.5 in.**

Calculate the 80% confidence interval for the population standard deviation.

- ~~1.224 < σ < 2.521~~
- **1.145 < σ < 2.358**
- ~~1.086 < σ < 2.124~~
- ~~1.310 < σ < 5.559~~

Ok, let's see what we know after reading the problem statement:

- **n = 8,**
- **s = 1.5in,**
- **α = 0.20,**
- **x-bar = 16.5in**

First we must find our critical chi-squared values with the NIST Chi-Squared Table associated with our alpha risk, sample size (8), degrees of freedom (7):

$$\textit{Confidence Interval for Standard Deviation}: \sqrt{\frac{(n-1)s^2}{X^2_{1-\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)s^2}{X^2_{\alpha/2}}}$$

$$X^2_{\alpha/2} = X^2_{.20/2} = X^2_{.10}$$

$$X^2_{1-\alpha/2} = X^2_{1-.20/2} = X^2_{1-.10} = X^2_{.90}$$

$$X^2_{.90,7} = 12.017 \qquad \& \qquad X^2_{.10,7} = 2.833$$

Now we can complete the equation using these chi-squared values along with the sample size, and sample standard deviation to calculate our interval.

$$\textit{Confidence Interval for Standard Deviation}: \sqrt{\frac{(n-1)s^2}{X^2_{1-\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)s^2}{X^2_{\alpha/2}}}$$

$$\sqrt{\frac{(8-1)1.5^2}{12.017}} < \sigma < \sqrt{\frac{(8-1)1.5^2}{2.833}}$$

$$1.145 < \sigma < 2.358$$

9. You manufacture a widget whose average length is 4.20 inches. You've upgraded your manufacturing equipment and you believe that it will not impact the overall length of the part.

You know the population standard deviation is 0.10 inches, and the sample mean of the 40 parts you measured is 4.24 inches. Using a 5% significance level to determine if the average length of the part has changed. Assume the length of the part is normally distributed.

Identify all of the statements below that are true:

- The null hypothesis, H_0: μ = ~~4.24~~ 4.20 inches **(False)**
- **The alternative hypothesis, H_a: μ ≠ 4.20 inches (True)**
- The hypothesis test is a ~~1-sided~~ two sided test **(False)**
- The critical rejection is ~~t_{crit}~~ z_{crit}= 1.96 **(False)**
- **The test statistic is z_{stat} = 2.53 (True)**
- The result of the test is the ~~failure to reject~~ rejection of the null hypothesis **(False)**

Because the problem statement is asking if the upgrade will "not impact the overall length", we can infer that this is a two-sided hypothesis test where the null and alternative hypothesis look like this:

H_0: μ = 4.20 inches H_a: μ ≠ 4.20 inches

Because we know population standard deviation and we're sampling more than 30 units, we can use the normal distribution to determine the critical rejection region which can be found using the NIST Z-Table.

With a 2-sided test at 5% significance we're looking for the Z-score that captures 47.5% of the area of the distribution. This is at **Z = 1.96 and Z = -1.96**.

We can now calculate the Test Statistic from the Sample Data:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{4.24 - 4.20}{\frac{0.10}{\sqrt{40}}} = 2.53$$

Now we can compare the z_{stat} (2.53) against the Rejection Criteria (-1.960, 1.960) and conclude that our test statistic is greater than our rejection criteria, therefore, we must reject the null hypothesis, in favor of the alternative hypothesis.

This question only had 2 true statements:

- **The alternative hypothesis, H_a: μ ≠ 4.20 inches (True)**
- **The test statistic is z_{stat} = 2.53 (True)**

The remaining statements were false, and I've shown the correct information above.

10. The one-way ANOVA Analysis below has 10 treatment groups with the total degrees of freedom of 19.

Variation Source	Sum of Squares (SS)	Degrees of freedom (DF)	Mean Squares (MS)	F-Value
Treatment (Between)		9		
Error (Within)	55			
Total	100	19		

Calculate the Treatment Mean Square for this ANOVA Table.

- 4.5
- **5**
- 5.5
- 6.1

First, we can solve for the treatment sum of squares by simply subtracting 55 from 100, to get a treatment sum of Square of **45**.

Then we must solve for the degrees of freedom.

The treatment degrees of freedom is equal to the number of treatment levels (10) – 1 = 9 degrees of freedom.

Then we can solve for the error degrees of freedom by subtracting 19 - 9; so, 10 degrees of freedom.

Now we can solve for the Treatment Mean Square:

$$\boldsymbol{Treatment\ Mean\ Square} = \frac{Sum\ of\ Squares\ of\ the\ Treatment}{Degrees\ of\ Freedom\ of\ the\ Treatment} = \frac{45}{9} = 5$$

Variation Source	Sum of Squares (SS)	Degrees of freedom (DF)	Mean Squares (MS)	F-Value
Treatment (Between)	**45**	**9**	5	**0.90**
Error (Within)	55	10	5.5	
Total	100	19		

11. You're creating a linear regression model for your data and you've calculated the following values.

S_{yy} = 31, S_{xy} = 12, S_{xx} = 76

Based on these results, what percentage of variation in Y, can be explained by the variation in X.

- 100%
- 81%
- 76%
- 43%
- 28%
- 13%
- **6%**
- 1%
- Not Enough Information Provided

This question is asking you to solve for the **Coefficient of Determination, R^2**, which **reflects the percentage of variation in Y that can be explained by the variation in X**.

Below is the equation to solve for the Pearson Correlation Coefficient, which then can be squared to find the coefficient of determination.

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}} * \sqrt{S_{yy}}} = \frac{12}{\sqrt{76} * \sqrt{31}} = 0.25$$

$$\boldsymbol{R^2 = r_{xy}{}^2 = 0.25^2 = 0.06 = 6\%}$$

12. You're creating a linear regression model for your data and you've calculated the following values.

S_{yy} = 4130, S_{xy} = 1527, S_{xx} = 626.86, β_0 = 17.81

What is the predicted value of Y when X = 23.

- 48
- 54
- 56
- 66
- **74**
- Not Enough Information Provided

Below is the regression equation to solve for Y.

$$Y(x) = \beta_0 + \beta_1 * x$$

To solve for Y, we need to calculate the slope coefficient:

$$\beta_1 = \frac{S_{xy}}{S_{xx}} = \frac{1527}{626.86} = 2.44$$

Now we can solve for **Y when X = 23**.

$$Y(23) = \beta_0 + \beta_1 * x = 17.81 + 2.44 * 23 = 73.84$$

13. **You're creating a linear regression model for your data and you've calculated the following values.**

S_{yy} = 31, S_{xy} = 12, S_{xx} = 76

What is correlation coefficient for this data set:

- -1
- -0.4
- 0
- 0.25
- 0.72
- 0.95
- 1
- Not Enough Information Provided

Below is the equation to solve for the **Pearson Correlation Coefficient**.

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}} * \sqrt{S_{yy}}}$$

Where

S_{yy} = 31
S_{xy} = 12
S_{xx} = 76

$$r_{xy} = \frac{12}{\sqrt{76} * \sqrt{31}} = 0.25$$

14. Which probability distribution is used to construct the c-chart:

- The Normal Distribution
- The Exponential Distribution
- **The Poisson Distribution**
- The Binomial Distribution

The c Charts utilize the **Poisson distribution** because it trends the number of defects associated with a process. Because it's possible for each item inspected to contain multiple defects, the Poisson distribution is required.

15. What type of control chart would be used to monitor the number of defectives for a process with a constant sample size:

- P Chart
- **NP Chart**
- C Chart
- U Chart

See the below Matrix which tells us that when monitoring the number of defectives for a process with a constant sample size, the **NP chart** is the correct control chart.

		Sample Size	
		Constant	Variable
Type	Defect	c Chart	u Chart
	Defectives	np Chart	p Chart

16. The Fake News Media Corporate collected data on the number of fake news stories published every day and constructed a p-chart.

A sample of 100 articles are inspected every day, however this can vary. The average percentage of fake news stories (defectives) was calculated as 0.106.

On a particular day, 200 articles were inspected and 47 fake news reports were observed. What is the conclusion of this day:

- The sample is in statistical control and this is a normal level of fake news
- **The sample is out of statistical control and there's a lot of fake news going around**

First, we must calculate the percentage of defectives found in the inspection.

We found 47 defectives (Fake news articles) in a sample of 200. Thus, our percentage defective for this sample is 0.235.

Now we must use our historical data to calculate the control limits for this process:

Where the historical average percentage of fake news stories (defectives) was calculated as 0.106.

$$\bar{p} = 0.106$$

$$UCL_{\bar{p}} = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{\bar{n}}} = 0.106 + 3\sqrt{\frac{0.106(1-0.106)}{200}} = 0.1713$$

$$LCL_{\bar{p}} = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{\bar{n}}} = 0.106 - 3\sqrt{\frac{0.106(1-0.106)}{200}} = 0.041$$

Our measured percentage defective (0.235) is greater than the upper control limit (0.1713), thus we can confirm that on this day, **the process is out of control and there is a lot of fake news going around.**

17. A p-chart monitors what type of attribute:

- The number of defective items in a subgroup
- The number of defects in a subgroup
- The percentage of defects in a subgroup
- **The percentage of defectives in a subgroup**

We can use the same attribute control chart matrix as before, where we can see that a **p-chart** is used to monitor processes with a variable samples size where defective items are the primary concern.

The percentage of defectives is monitored because the sample size is variable; thus a p-chart could be described as monitoring **the percentage of defectives in a subgroup**

		Sample Size	
		Constant	**Variable**
Type	**Defect**	c Chart	u Chart
	Defectives	np Chart	p Chart

18. What is the UCL for a p-chart when the average daily inspection quantity is 50, and the historical percentage of defectives is 0.05?

- 0.21
- 0.09
- 0.29
- **0.14**
- 0.17

The **upper control limit** for **a P-chart** can be calculated using the following equation:

$$UCL_{\bar{p}} = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{\bar{n}}}$$

The problem statement also gives us the variables needed to solve this equation:

- $\bar{n} = 50$
- $\bar{p} = 0.05$

$$UCL_{\bar{p}} = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{\bar{n}}} = 0.05 + 3\sqrt{\frac{0.05(1-0.05)}{50}}$$

$$UCL_{\bar{p}} = 0.05 + 3\sqrt{0.00095} = 0.142$$

19. You're constructing an NP chart, where you've sampled from 25 subgroups, each with 100 samples, and found a total of 145 defective units.

Calculate the UCL for this process.

- Not Enough Information Provided
- 5.8
- 0.058
- 7.0
- **12.8**
- 14.5

The upper control limit of an NP Chart is calculated using the following equation:

$$UCL_{np} = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$$

To execute this equation, we need to know the following variables:

$$\bar{p} = \%\, Defective$$

$$n\bar{p}\ Centerline = \frac{\sum np}{k}$$

Let's use our information from the problem statement to calculate these variables:

$$\bar{p} = \frac{\sum np}{\sum n} = \frac{Sum\ of\ All\ Defectives}{Sum\ of\ Subgroup\ Quantity} = \frac{145}{2500} = 0.058$$

$$n\bar{p}\ Centerline = \frac{\sum np}{k} = \frac{Sum\ of\ All\ Defectives}{\#\ of\ subgroups} = \frac{145}{25} = 5.8$$

Now we can plug these variables back into the equation for the upper control limit:

$$UCL_{np} = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$$

$$UCL_{np} = 5.8 + 3\sqrt{5.8(1-0.058)} = 5.8 + 7.0 = 12.8\text{jmn}$$

$$\boldsymbol{UCL_{np} = 5.8 + 7.0 = 12.8}$$

20. **You're manufacturing a widget and using an X-bar and R chart to control the critical feature of the product. Your normal process has the following attributes:**

X-double bar is 225, R-bar is 12, n = 8.

Identify the upper and lower control limits for the X-bar chart:

- 0
- **220.52**
- **229.48**
- 1.63
- 233.14
- 218.71
- 22.37

Below are the equations for the control limits of an X-bar and R chart:

$$\boldsymbol{Upper\ Control\ Limit}: \boldsymbol{UCL_{\bar{X}}} = \bar{\bar{X}} + A_2\bar{R}$$

$$\boldsymbol{Lower\ Control\ Limit}: \boldsymbol{LCL_{\bar{X}}} = \bar{\bar{X}} - A_2\bar{R}$$

Where:

- $\bar{\bar{X}} = 225$
- $\bar{R} = 12$

Now we must look up the A_2 constant using the sample size (n=8), and we find A_2 = 0.373

With this we can now calculate the upper and lower control limits:

$$\boldsymbol{UCL_{\bar{X}}} = \bar{\bar{X}} + A_2\bar{R}$$

$$\boldsymbol{UCL_{\bar{X}}} = 225 + 0.373 * 12 = 229.48$$

$$\boldsymbol{LCL_{\bar{X}}} = \bar{\bar{X}} - A_2\bar{R}$$

$$\boldsymbol{LCL_{\bar{X}}} = 225 - 0.373 * 12 = 220.52$$

X-Bar and R Chart				
Subgroup Sample Size	X-Bar Factor	Range Factors		Variance Factor
n	A_2	D_3	D_4	d_2
2	1.880	-	3.267	1.128
3	1.023	-	2.575	1.693
4	0.729	-	2.282	2.059
5	0.577	-	2.115	2.326
6	0.483	-	2.004	2.534
7	0.419	0.076	1.924	2.704
8	0.373	0.136	1.864	2.847
9	0.337	0.184	1.816	2.970
10	0.308	0.223	1.777	3.078
15	0.223	0.347	1.653	3.472
20	0.180	0.415	1.585	3.735
25	0.153	0.459	1.541	3.931

21. Calculate Cpk for the following Parameters: (USL = 1.35, LSL = 1.15, σ = 0.025, μ = 1.25)

- 0.67
- 1.0
- **1.33**
- 1.67
- 2.0

Below is the equation for C_{pk}.

$$C_{pk} = Min(C_{p,Lower}, C_{p,Upper}) = Min\left(\frac{USL - \mu}{3s}, \frac{\mu - LSL}{3s}\right)$$

Where we know the following variables:

- **USL = 1.35,**
- **LSL = 1.15,**
- **σ = 0.025,**
- **μ = 1.25**

Let's plug these in and see what we come up with:

$$C_{pk} = Min\left(\frac{1.35 - 1.25}{3 * 0.025}, \frac{1.25 - 1.15}{3 * 0.025}\right)$$

$$C_{pk} = Min\left(\frac{0.10}{0.075}, \frac{0.10}{0.075}\right)$$

$$\boldsymbol{C_{pk} = Min(1.33, 1.33) = 1.33}$$

22. Calculate Cpk for the following Parameters: (USL = 205, LSL = 145, σ = 10, μ = 190)

- **0.50**
- 0.67
- 1.0
- 1.33
- 1.50

Below is the equation for C_{pk}.

$$C_{pk} = Min(C_{p,Lower}, C_{p,Upper}) = Min\left(\frac{USL - \mu}{3s}, \frac{\mu - LSL}{3s}\right)$$

Where we know the following variables:

- **USL = 205,**
- **LSL = 145,**
- **σ = 10,**
- **μ = 190**

Let's plug these in and see what we come up with:

$$C_{pk} = Min\left(\frac{205 - 190}{3 * 10}, \frac{190 - 145}{3 * 10}\right)$$

$$C_{pk} = Min\left(\frac{15}{30}, \frac{45}{30}\right)$$

$$\boldsymbol{C_{pk} = Min(0.5, 1.5) = 0.5}$$

23. What Cpk value will theoretically result in 1 defect per million?

- 1.0
- 1.33
- **1.66**
- 2.0
- 6.0

Using the table below you can take your C_{pk} value and cross reference the % Defect, Defects Per Million or Defect Rate - and this value can become the Likelihood of Failure within your FMEA.

You can see that a C_{pk} of 1.66 should result in a DPM (defects per million) of 1.

C_{pk}	Sigma Level (σ)	% Conforming	Defects Per Million (DPM)	Defect Rate
.33	1	68.27%	317,311	1 in 3.15
.67	2	95.45%	45,500	1 in 22
1.00	3	99.73%	2,700	1 in 370
1.33	4	99.99%	63	1 in 15,873
1.66	5	99.9999%	1	1 in 1,000,000
2.00	6	99.9999998%	.002	1 in 500,000,000

24. How many treatments would be required for a DOE with 8 factors where a quarter factorial design is chosen:

- 256
- 128
- **64**
- 32
- 16
- 8

Calculating the number of treatments for a quarter factorial design can be calculated as such:

$$\boldsymbol{Quarter}\ Factorial\ Design: Number\ of\ Treatments = \frac{Levels^{Factors}}{4}$$

$$\boldsymbol{Quarter\ Factorial\ Design: Number\ of\ Treatments = \frac{L^F}{4} = \frac{2^F}{2^2} = 2^{8-2} = 2^6 = 64}$$

25. **You performed a full factorial DOE to improve the yield of a process with two factors at two levels and have measured the following response values.**

What is the estimated effect of Factor B:

		Factors		Response
		A	B	% Yield
Treatments	1	+	+	64
	2	-	+	75
	3	+	-	87
	4	-	-	95

- -9.5
- **-21.5**
- 11
- -1.5
- -8

Below is the calculation for estimating the effect of a factor:

$$\boldsymbol{Estimated\ Effect = Average\ at\ High - Average\ at\ Low}$$

Factor B is at the "High" (+) level in treatments 1 & 2, and at the "Low" (-) level in treatments 3 & 4. Let's plug in the response values for these treatments to estimate the effect of Factor B.

$$\boldsymbol{Factor\ A\ Estimated\ Effect = \frac{64+75}{2} - \frac{87+95}{2} = -21.5}$$