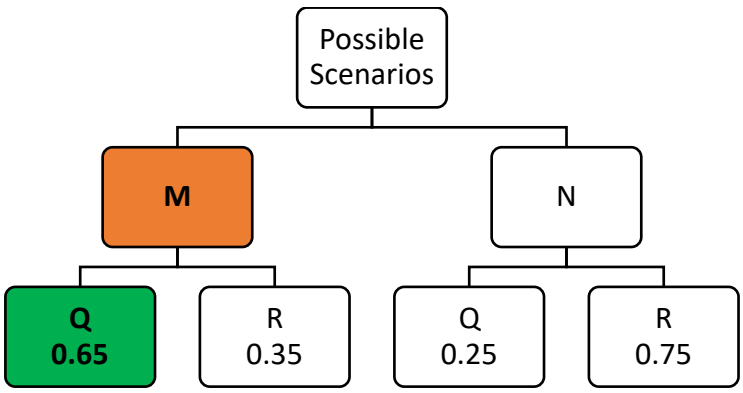
1. **Different data types provide different amount of information. What is the correct order of data types providing the most to the least information?**
   A) Interval > Ratio > Ordinal > Nominal
   B) Ordinal > Ratio > Nominal > Interval
   C) Ratio > Interval > Ordinal > Nominal
   D) Nominal > Ordinal > Interval > Ratio

2. **There are 140 pens in a box. 11 of them are black, 8 of them are green, 15 of them are red, and the rest of them are all blue pens. You randomly take out a pen from the box to take some quick notes. What is the probability of it being a blue pen?**
   A) 24%
   B) 76%
   C) 32%
   D) 100%

$$Number\ of\ blue\ pens: \quad 140 - (11 + 8 + 15) = 140 - 34 = 106$$

$$Probability\ of\ taking\ a\ blue\ pen: \quad \boldsymbol{P(Blue)} = \frac{106}{140} = 0.76 = \mathbf{76\%}$$

3. **The following diagram depicts the probability of two possible scenarios M and N. What is the probability that event Q will take place considering that event M has occurred?**
   A) 0.90
   B) 0.45
   C) 0.25
   D) 0.65

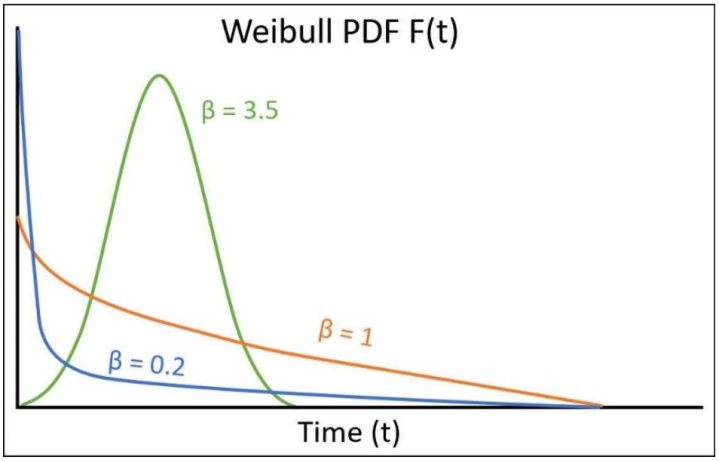


Because event M has already occurred, the probability of **Q** is **0.65** as in the diagram.

4. The _____ distribution is able to model variety of shapes and datasets.

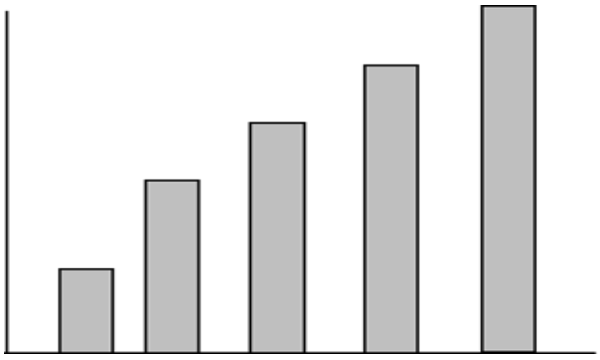
A) Normal
B) Poisson
C) Exponential
D) Weibull

By changing the slope parameter (β) of the **Weibull distribution**, it can model variety of shapes.



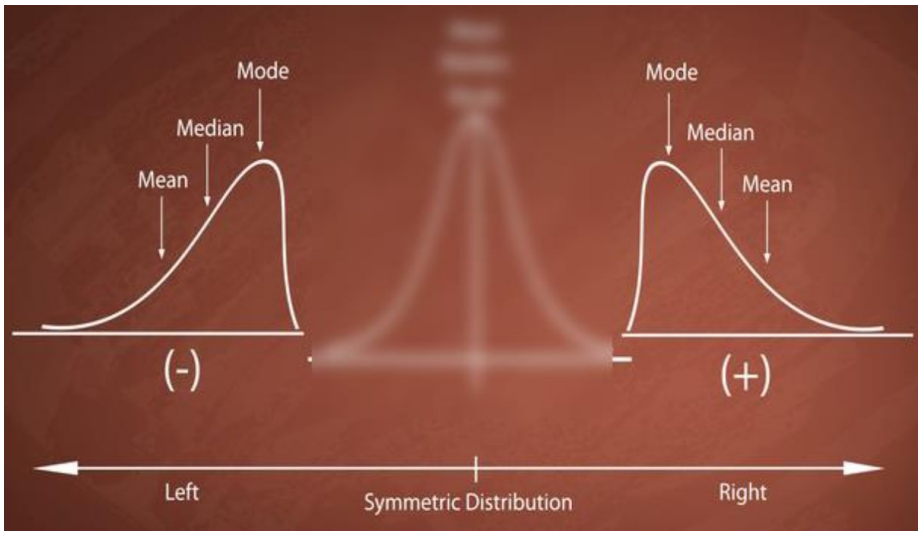
5. The following histogram shows the distribution of a dataset. What is the relative situation of mode, median and mean according to this histogram?

A) Mode > Median > Mean
B) Mode < Mean = Median
C) Median < Mode = Mean
D) Mean > Median > Mode



This histogram represents a dataset with a negative skew (as in the left picture below). In this situation, the mean value is the one most affected by outliers, while the mode is nearly unaffected. Hence:

Mode > Median > Mean



6. **You are producing a product that has customers worldwide. According to the recorded data, it has an average of 0.8 complaints per hour. What is the probability that you get zero complaints in an hour?**

   A) 45%
   B) 0%
   C) 2%
   D) 36%

   Because the problem statement gives us the *average number of complaints per hour*, this is lambda, the failure rate associated with the Poisson distribution. To calculate the probability of occurrence when using the Poisson distribution, we use the following equation:

$$f(x) = P(X = x) = \frac{e^{-\lambda} * \lambda^{x}}{x!}$$

$$f(0) = P(X = 0) = \frac{e^{-0.8} * 0.8^{0}}{0!} = 0.45 = \mathbf{45\%}$$

7. **A system is in its useful life period and has shown to have an MTBF of 800 hours. What is the reliability of the system at 400 hours?**

   A) 50%
   B) 25%
   C) 61%
   D) 2%

$$Reliability: \quad R(t) = e^{-\lambda t} = e^{\frac{-t}{\theta}}$$

$$\boldsymbol{R(400)} = e^{\frac{-t}{\theta}} = e^{\frac{-400}{800}} = e^{-0.5} = 0.61 = \mathbf{61\%}$$

8. **When using the binomial distribution, what is the correct formula for the probability function?**

   A) $f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$ — Poisson Distribution

   B) $f(x) = \binom{n}{x} p^{x}(1-p)^{n-x}$ — **Binomial Distribution**

   C) $f(t) = \frac{\beta}{\theta}(\frac{t}{\theta})^{\beta} e^{-(\frac{t}{\theta})^{\beta}}$ — Weibull Distribution (PDF)

   D) $f(t) = 1 - e^{-(\frac{t}{\theta})^{\beta}}$ — Weibull Distribution (CDF)

9. **The standard deviation of a population is 36. We take 20 samples from this population. What's the standard deviation of sample means?**

   A) 3.33
   B) 1.80
   C) 36.0
   D) 8.05

$$\boldsymbol{Standard\ deviation\ of\ sample\ means} = \frac{\sigma}{\sqrt{n}} = \frac{36}{\sqrt{20}} = \mathbf{8.05}$$

10. You have a dataset with normal distribution which has a mean of 23 and standard deviation of 3.2. Your LSL is 18. What portion of the results are in the accepted range?

A) 94%

B) 44%

C) 6%

D) Not Enough Information

There is no USL specified in the question. So, all the results higher than LSL are in the accepted range.

$$Accepted\ Results:\ from\ LSL\ to\ +\infty\ = (LSL\ to\ Mean) + (Mean\ to\ +\infty)$$

First, we have to calculate the Z score associated with the LSL (18):

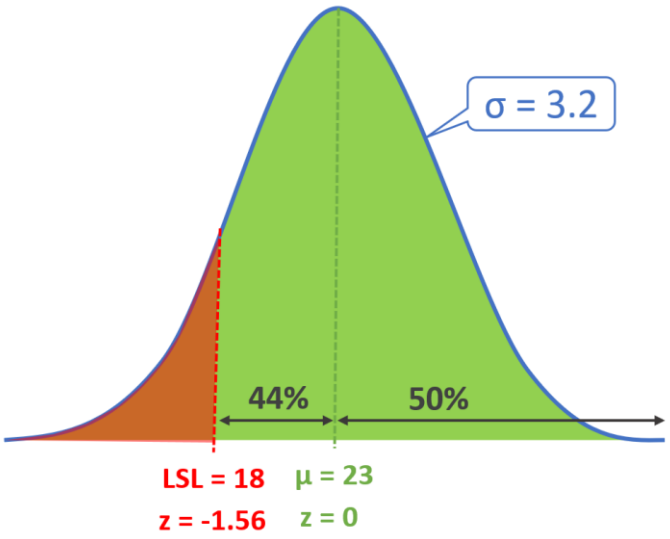
$$Z = \frac{X - \mu}{\sigma} = \frac{18 - 23}{3.2} = -1.56$$

So, 0.44 or 44% of the results are between LSL and the mean.

Also, 50% of the results are in the right side of the mean.

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	0.41309	0.41466
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922
1.5							0.44062	0.44179
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254



$$Accepted\ Results:\ from\ LSL\ to\ +\infty\ = (LSL\ to\ Mean) + (Mean\ to\ +\infty)$$

$$\boldsymbol{Accepted\ Results} = 44\% + 50\% = 94\%$$

11. Which of the following items increases the power of hypothesis tests?

A) Decreasing the alpha risk
B) Increasing sample size
C) Using a two-sided hypothesis test
D) All of the above

12. While performing an ANOVA Analysis, what is the null hypothesis?

A) Sample means are different
B) Sample variances are equal
C) Sample variances are different
D) Sample means are equal

In an ANOVA analysis, the null hypothesis is always this: "**Sample means are equal**".

13. For the data variables given in the following table, calculate the coefficient of correlation.

A) 1.00
B) 1.15
C) 0.87
D) 0.67

We expand the table to calculate the necessary items:

X	Y	X^2	Y^2	X*Y
2	1	4	1	2
4	3	16	9	12
7	6	49	36	42
8	7	64	49	56
12	10	144	100	120
$\sum X_i = 33$	$\sum Y_i = 27$	$\sum X_i^2 = 277$	$\sum Y_i^2 = 195$	$\sum (X_i * Y_i) = 232$

Now we can use the above sums to calculate S_{xy} , S_{xx} and S_{yy} :

$$S_{xy} = \sum (X_i * Y_i) - \frac{(\sum X_i) * (\sum Y_i)}{n} = 232 - \frac{33 * 27}{5} = 53.8$$

$$S_{xx} = \sum X_i^2 - \frac{(\sum X_i)^2}{n} = 277 - \frac{33^2}{5} = 59.2$$

$$S_{yy} = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = 195 - \frac{27^2}{5} = 49.2$$

Finally, we can calculate the coefficient of correlation:

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}} * \sqrt{S_{yy}}} = \frac{53.8}{\sqrt{59.2} * \sqrt{49.2}} = \mathbf{1.00}$$

14. What type of control chart is used for controlling defectives when the sample size is variable?

A) NP Chart

B) P Chart

C) C Chart

D) U Chart

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

15. We have the following information about a process. What is the lower control limit for the X-bar?

$$\bar{\bar{X}} = 40, \bar{s} = 10, n = 6, \sum s = 80, \sum \bar{X} = 320$$

A) 14.27

B) 27.13

C) 25.73

D) 12.87

We should use the X-bar and S chart, because the question has provided S data.

X-Bar and S Chart				
Subgroup Sample Size	X-Bar Factor	Standard Deviation Factors		Variance Factor
n	A_3	B_3	B_4	c_4
2	2.659	-	3.267	0.7979
3	1.954	-	2.568	0.8862
4	1.628	-	2.266	0.9213
5	1.427	-	2.089	0.9400
6	1.287	0.030	1.970	0.9515
7	1.182	0.118	1.882	0.9594
8	1.099	0.185	1.815	0.9650
9	1.032	0.239	1.761	0.9693
10	0.975	0.284	1.716	0.9727
15	0.789	0.428	1.572	0.9823
20	0.680	0.510	1.490	0.9869
25	0.606	0.565	1.435	0.9896

$$Lower\ Control\ Limit\ for\ \bar{X}: \quad LCL_{\bar{X}} = \bar{\bar{X}} - A_3\bar{s}$$

$$LCL_{\bar{X}} = 40 - 1.287 * 10 = \mathbf{27.13}$$

16. Which of the following conditions are true when C_{pk} equals C_p?

A) $\frac{USL-LSL}{3\sigma} = \bar{\bar{X}}$

B) $\frac{USL+\bar{\bar{X}}}{\sigma} = LSL$

C) $USL - LSL = \bar{\bar{X}}$

D) $\frac{USL+LSL}{2} = \bar{\bar{X}}$

When C_{pk} equals C_p, this means that ***the process is centered around the mean***.
Hence, we can say that the mean is in the center of the upper and lower specification limits:

$$\frac{USL + LSL}{2} = \bar{\bar{X}}$$

17. We are performing a 2^4 DOE, with one replicate measurement per treatment. What is the total degrees of freedom of this DOE?

A) 4

B) 7

C) 8

D) 15

$$DOE: \quad Levels^{Factors} = 2^4$$

$$Levels = 2 \quad ; \quad Factors = 4$$

Ok, so in this DOE, we have 4 factors with 2 levels each.

Apparently, this is a **full-factorial DOE**. Hence:

$$Number\ of\ Treatments: \quad Levels^{Factors} = 2^4 = 16$$

So, we have **16 treatments (N)** in this DOE.

The total degrees of freedom in a DOE is equal to N-1, so here we have 16 – 1 = **15** degrees of freedom.

The simplified ANOVA table of such DOE would look like the following:

Variation Source	Degrees of Freedom (DF)
Factor A	1
Factor B	1
Factor C	1
Factor D	1
Error	11
TOTAL	**15**

(Scroll down to the next page to see an alternative method to solve this question.)

Well, another way to solve this problem, is using the following equation:

$$\boldsymbol{Total\ Degrees\ of\ Freedom} = \boldsymbol{N} - \mathbf{1}$$

$$\boldsymbol{N} = \boldsymbol{L_A} \times \boldsymbol{L_B} \times \boldsymbol{L_C} \times \boldsymbol{L_D} \times \boldsymbol{n}$$

Where:

L_A: $The\ number\ of\ levels\ of\ Factor\ A$

L_B: $The\ number\ of\ levels\ of\ Factor\ B$

L_C: $The\ number\ of\ levels\ of\ Factor\ C$

L_D: $The\ number\ of\ levels\ of\ Factor\ D$

n: $The\ number\ of\ replicates\ per\ treatmen\ \ (by\ default:\ n = 1)$

So, in our example, we have 2 levels for all the 4 factors. Hence:

$$N = 2 \times 2 \times 2 \times 2 \times 1 = 16$$

$$\boldsymbol{Total\ Degrees\ of\ Freedom} = 16 - 1 = \mathbf{15}$$

A DOE table with 4 factors with 2 levels each, would look like the following:

As it's illustrated in the table, there are 16 treatments in this full-factorial DOE.

	Temperature	Time	Eggs	Sugar
Treatment 1	+	+	+	+
Treatment 2	+	+	+	-
Treatment 3	+	+	-	+
Treatment 4	+	+	-	-
Treatment 5	+	-	+	+
Treatment 6	+	-	+	-
Treatment 7	+	-	-	+
Treatment 8	+	-	-	-
Treatment 9	-	+	+	+
Treatment 10	-	+	+	-
Treatment 11	-	+	-	+
Treatment 12	-	+	-	-
Treatment 13	-	-	+	+
Treatment 14	-	-	+	-
Treatment 15	-	-	-	+
Treatment 16	-	-	-	-

18. The number of treatments in a DOE equals 2^{6-2}, and you have 6 factors. What kind of DOE is this?

A) ¼ Factorial

B) ½ Factorial

C) Full Factorial

D) Infinite

$$Full\ Factorial\ Design:\ \ Number\ of\ Treatments = Levels^{Factors} = 2^6$$

$$^1/_2\ Factorial\ Design:\ \ Number\ of\ Treatments = \frac{Levels^{Factors}}{2} = \frac{2^6}{2} = 2^{6-1}$$

$$^1/_4\ \boldsymbol{Factorial}\ Design:\ \ Number\ of\ Treatments = \frac{Levels^{Factors}}{4} = \frac{2^6}{4} = 2^{6-2}$$

The number of treatments matches the **¼ Factorial** Design formula.

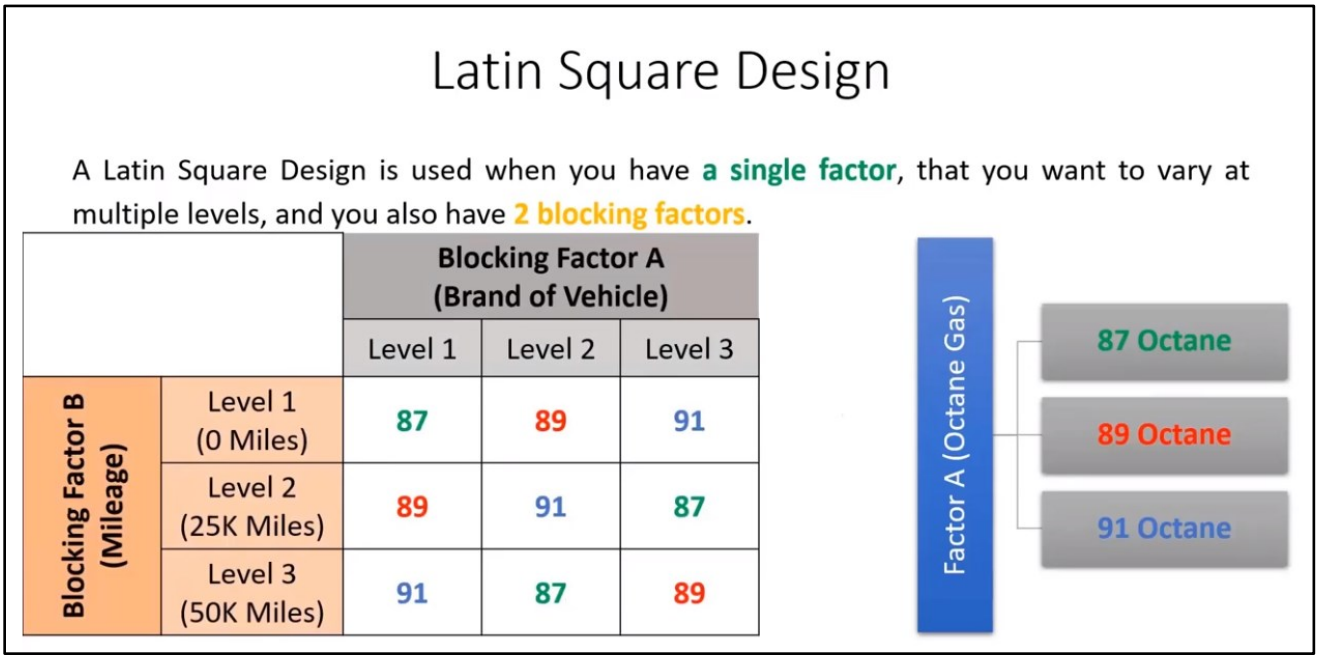
19. What is the disadvantage of using Latin Square method in DOE?

A) Not being able to study interactions

B) Requires more samples than a full factorial experiment

C) Does not allow for blocking

D) The results are often confounding



20. The air pressure is used in a process to change a critical dimension of the product. If we perform a DOE for this process, what is the air pressure considered to be?

A) Response

B) Dependent

C) Independent

D) Covariant

In this process, the air pressure is an input:

Input = Factor = **Independent** Variable

Output = Response = Dependent Variable

21. You are planning to perform a DOE. How would you avoid confounding?

A) Increase blocking

B) Increase randomization

C) Increase variables

D) Plan for a fractional factorial DOE

22. Employees in an investment organization store documents both as scanned files in computers and also as hard copies in filing cabinets. Which type of waste is this?

A) Extra Processing
B) Over-Production
C) Inventory
D) Motion

Extra production is the waste that occurs when you put in more effort, time or work into a production step than what is needed. In this instance, we are performing double-work to store an item, when one would suffice.

23. Delivering goods directly to the _____ is called just-in-time.

A) Customer
B) Supply Chain Department
C) Engineering Department
D) Manufacturing Line

24. What is the correct order for a design process?

A) Design > Validation > Review > Verification
B) Inputs > Design > Verification > Validation
C) Review > Validation > Design > Verification
D) Inputs > Validation > Verification > Review