Question 1 part A

What assumptions do you need to make, given the information?

You need to assume that each of the three events (i.e. completing the project on time) are Independent Events.

What is the probability of finishing all three events on time

P(X)xP(Y)xP(Z) = 0.99 x 0.95 x 0.80

What is the probability that only projects X and Y are finished on time.

A fuller description of the scenario would include a statement that Z had not finished on time.

P(X) xP(Y)xP(Zc) = 0.99 x 0.95 x 0.20

Question 1 Part B

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | CompSci | Stats | Equine Sci | Total |
| Males | 140 | 100 | 20 | 260 |
| Females | 30 | 80 | 30 | 140 |
| Total | 170 | 180 | 50 | 400 |

P(CompSci) = 170/400 = 42.5%

P(Female and Stats) = 80/400 = 20%

P(Male) = 260/400 = 65%

P(Female|Stats) = 80/180 = 44.44%

P(Male or Stats) = P(Male) + P(Stats) – P(Male and Stats) = (260/400) + (180/400) – (100/400) = 340/400 = 85%

P(Eq.Science|Female) = 30/140 =

Question 2A

|  |
| --- |
| > dpois(5:15,lambda=10)  [1] 0.03783327 0.06305546 0.09007923 0.11259903 0.12511004 0.12511004  [7] 0.11373640 0.09478033 0.07290795 0.05207710 0.03471807  >  > ppois(5:15,lambda=10)  [1] 0.06708596 0.13014142 0.22022065 0.33281968 0.45792971 0.58303975  [7] 0.69677615 0.79155648 0.86446442 0.91654153 0.95125960  >  > pbinom(5:15,size=100,prob=0.10)  [1] 0.05757689 0.11715562 0.20605086 0.32087389 0.45129017 0.58315551  [7] 0.70303310 0.80182111 0.87612321 0.92742703 0.96010947  >  > dbinom(5:15,size=100,prob=0.10)  [1] 0.03386580 0.05957873 0.08889525 0.11482303 0.13041628 0.13186535  [7] 0.11987759 0.09878801 0.07430209 0.05130383 0.03268244 |

|  |
| --- |
| > 0:5  [1] 0 1 2 3 4 5  >  > ppois(0:5,lambda=2)  [1] 0.1353353 0.4060058 0.6766764 0.8571235 0.9473470 0.9834364  >  > ppois(0:5,lambda=(1/2))  [1] 0.6065307 0.9097960 0.9856123 0.9982484 0.9998279 0.9999858  >  > dpois(0:5,lambda=2)  [1] 0.13533528 0.27067057 0.27067057 0.18044704 0.09022352 0.03608941  >  > dpois(0:5,lambda=(1/2))  [1] 0.6065306597 0.3032653299 0.0758163325 0.0126360554 0.0015795069  [6] 0.0001579507  >  > pbinom(0:5,size=2,prob=0.05)  [1] 0.9025 0.9975 1.0000 1.0000 1.0000 1.0000 |

Question 3

Part A

All of the Probabilities must sum to 1.

Necessarily p= 0.13

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 |
| P(X) | 0.16 | 0.14 | 0.13 | 0.17 | 0.21 | 0.19 |

Compute E(X) and E(X2)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 |  | SUMS |
| P(X) | 0.16 | 0.14 | 0.13 | 0.17 | 0.21 | 0.19 |  | 1 |
|  |  |  |  |  |  |  |  |  |
| X.P(X) | 0.16 | 0.28 | 0.39 | 0.68 | 1.05 | 1.14 |  | 3.7 |
| X2.P(X) | 0.16 | 0.56 | 1.17 | 2.72 | 5.25 | 6.84 |  | 16.7 |

E(X) = 3.7 E(X2) = 16.7

***V(X) = E(X2) - E(X)2***= 16.7 – (3.7)2 = 16.7 – 13.69 = 3.01

|  |  |  |
| --- | --- | --- |
| Test Value | = | Z score |
| 325 | = | 1.5 |
| 300 | = | 1 |
| 150 | = | -2 |
| 250 | = | 0 |

1-

|  |
| --- |
| > Zs  [1] -2.25 -2.00 -1.75 -1.50  [5] -1.25 -1.00 -0.75 -0.50  [9] -0.25 0.00 0.25 0.50  [13] 0.75 1.00 1.25 1.50  [17] 1.75 2.00 2.25  >  > pnorm(Zs)  [1] 0.01222447 0.02275013 0.04005916 0.06680720  [5] 0.10564977 0.15865525 0.22662735 0.30853754  [9] 0.40129367 0.50000000 0.59870633 0.69146246  [13] 0.77337265 0.84134475 0.89435023 0.93319280  [17] 0.95994084 0.97724987 0.98777553 |

Shapiro Test

The null hypothesis is that the data is normally distributed

The alternative hypothesis is that the data set is not normally distributed

The QQ plot

This is a graphical procedure to complement formal tests of normality, such as the Shapiro Wilk Test. If the data is normally distributed, then the data points will follow the diagonal trendline more or less.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F | G | H |
| Before | 170 | 180 | 184 | 183 | 186 | 184 | 169 | 130 |
| After | 190 | 194 | 200 | 199 | 197 | 200 | 185 | 145 |
| Diff | 20 | 14 | 16 | 16 | 11 | 16 | 16 | 15 |
|  |  |  |  |  |  |  |  |  |
|  |  | 15.5 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | 400 | 196 | 256 | 256 | 121 | 256 | 256 | 225 |
|  |  |  |  |  |  |  |  |  |
|  |  | 1966 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | 6.285714 |  |  |  |  |  |  |  |

Hypotheis Testing process

|  |  |  |
| --- | --- | --- |
| 1 | Write out the Null hypothesis and alternative hypothesis | |
| 2 | Compute the the test statistic, using the appropriate standard error va;ue | |
| 3 | determine the critical value | determine the p-value |
| 4 | Decision Rule for TS and CV | Decision Rule for p-value and threshold |

One Sample tests

* 1 sample mean
* Single sample proportion
* Paired sample test (specifically the case-wise differences)

Two sample

* Difference of two independent sample means
* Difference in two sample proportions

Other tests

* Variance test
* Shapiro Wilk Test

Additional Considerations

1. Is the test one tailed or two tailed? (i.e. k=1 or 2?)
2. It will usually be two tailed. (Confidence Intervals are always two tailed)
3. Is the aggregrate sample size large or small? A large sample is composed of thirty or more observations.
4. |If the sample is small use the student t distribution , specifying the degrees of freedom
5. If the sample is large – use the Z distribution (qnorm())(qt())
6. If the procedure is two tailed, the values can be re-aranged to compute the confidence interval.

Computing the pooled variance (from formulae)

Computing the pooled variance (from formulae)

The null hypothesis and alternative hypothesis can be expressed as follows:

An alternative specificication of the null and laternative hypothesis , which states the expected alue under the null hypothesis is as follows

**The Shapiro-Wilk Test**

The Shapiro-Wilk Test is a formal hypothesis test for testing normality.

The data set is normally distributed

The data set is not normally distributed

|  |
| --- |
|  |

**Variance Testing**

The students are given the R code indicating the standard deviations for both data sets.

The null and alternative hypothesis can be formally specified in either tof the two following ways. The second way considers the problem in the context of variance ratio. If the variance e ratio is 1, then the variance s are equal.

This approach is more consistent with the required computation.

|  |  |
| --- | --- |
|  | : |

Computing the test statistic

The p-value will be provided in the R code output. Students are required to state a conclusion based on this p-value