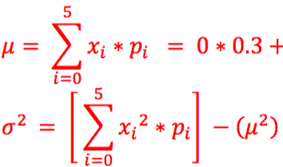
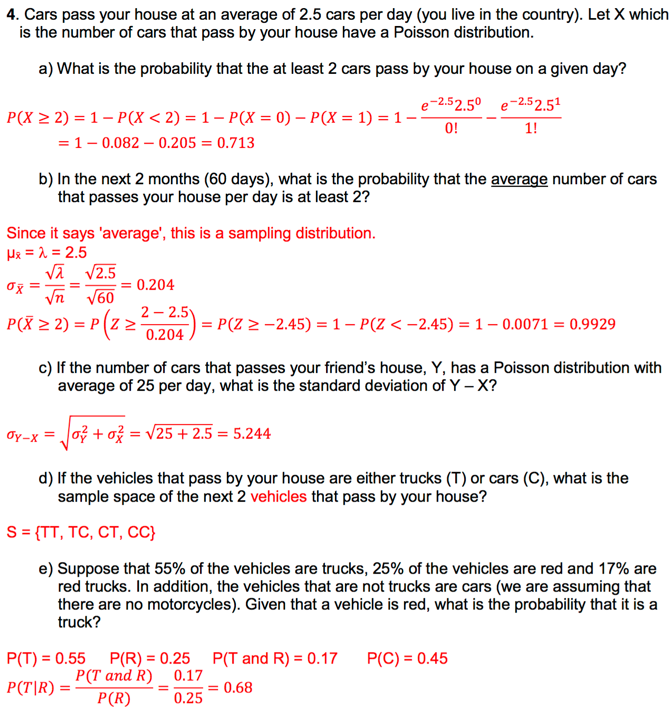


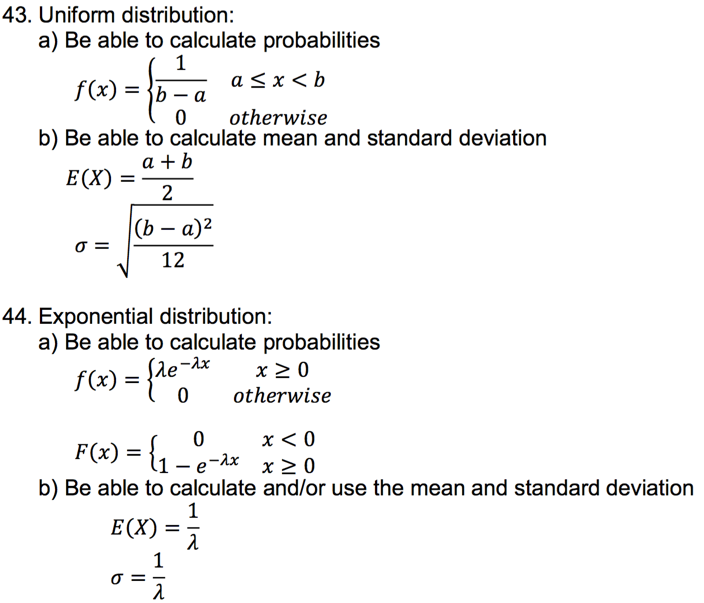
../Desktop/屏幕快照%202016-05-04%20下午11.52.16.png../Desktop/屏幕快照%202016-05-04%20下午11.46.30.png

**../Downloads/b8525313480be649bed306ba81018709.png**../Desktop/屏幕快照%202016-05-04%20下午11.39.00.png../Desktop/屏幕快照%202016-05-04%20下午11.39.28.png

Lower and high outliers:

../Desktop/屏幕快照%202016-05-04%20下午8.38.22.png

Min Q1 Median Q3 Maximum

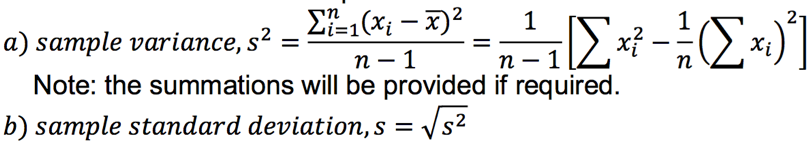


45. Poisson Distribution 

mean = variance =

如果单位时间发生的次数（如到达的人数）服从参数为r的泊松分布，则任连续发生的两次时间的间隔时间序列服从参数为r的指数分布

泊松分布是单位时间内独立事件发生次数的概率分布， 指数分布是独立事件的时间间隔的概率分布



experiment: control, randomization, replication

|  |  |
| --- | --- |
| Reject if | Fail to Reject if |
| The confidence interval does NOT contain the null value (or 0). | The confidence interval does contain the null value (or 0). |
| 𝑝≤𝛼 | 𝑝>𝛼 |
| Reject H0. There is evidence that the population mean (difference) is different from/greater than/less than \_\_\_\_\_\_\_\_. | Fail to Reject H0. There is NOT evidence that the population mean (difference) is different from /greater than/less than \_\_\_\_\_\_\_\_. |

Type I error: We reject H0 when H0 is true

P(Type I error) = α

Type II error: We fail to reject H0 when H0 is false P(Type II error)=β,Power=1-β

|  |  |  |
| --- | --- | --- |
|  | **Z-Test** | **T-test** |
| Hypotheses:  Null  Alternative:  Two-Sided  One Sided:  Right Tailed  Left Tailed | α/2  α | α/2  α |
| Test Statistic: |  |  |
| P-value: Two-sided  Right Tailed  Left Tailed | 2P(Z<z)  P(Z>z) = 1-P(Z<z)  P(Z<z) | 2P(T>t)  P(T>t) DF = n-1  P(T>|t|) |
| Confidence Interval: |  |  |

|  |  |
| --- | --- |
| **Two-Sample Comparison of Means T-test** | |
| Hypotheses: Null  Alter: Two side  One-sided Right  One-sided Left |  |
| Test Statistic: |  |
| Satterthwaite Degrees of Freedom |  |
| P-value Two-sided  Right Tailed  Left Tailed | 2P(T>t) Use T-table  P(T>t) DF=Satterthwaite  P(T>|t|) |
| Confidence Interval: |  |

|  |  |
| --- | --- |
| **Matched Pairs T-test** | |
| Hypotheses: Null  Alter: Two side  One-sided Right  One-sided Left |  |
| Test Statistic: |  |
| P-value Two-sided  Right Tailed  Left Tailed | 2P(T>t) Use T-table  P(T>t) DF=n-1  P(T>|t|) |
| Confidence Interval: |  |

**前提**: normal or n>30

Take random sample -> reduce bias

Large sample size -> Less variability

Larger sample has a same center as smaller

Extraneous variables matched -> block test (ex: 40 female and 40 male mice divided in two groups with same number of sex drink different types of water : block gender)

No blocking -> satisfied

Statistics theta is unbiased if E(theta) = theta

Be able to calculate the sample size, n, needed for a particular margin of error, ME (half-width). is calculated for the preliminary study (the study that obtained s)

|  |  |
| --- | --- |
| one-sample z | one-sample t |
|  |  |

|  |  |
| --- | --- |
| lower bound | upper bound |
|  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | one-sample z | one-sample t | 2-sample z(t) independent | 2-sample t(z) pair |
| μ |  |  |  |  |
| df | N/A | n – 1 | will be given to you | n - 1 |

Interpret: We are \_\_ confident that the true mean of the \_\_\_\_\_\_ is in the interval (\_\_\_, \_\_\_\_).

The following are the test statistics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | one-sample | | two-sample | |
| independent | pairs |
| μ |  |  |  |  |
| df | N/A | n - 1 | will be given to you | n - 1 |

Be able to use the 4 steps in hypothesis testing:

1. Identify the parameter(s) of interest and describe it (them) in the context of the problem situation.
2. State the Hypotheses.
3. Calculate the appropriate test statistic, df (if necessary), find the P-value (for z tests, you will need to be able to calculate this, for t tests, the value will be provided.)
4. Decision, reason for your decision including numbers, conclusion in the problem context (see SLIDEs for the wording).

|  |  |  |  |
| --- | --- | --- | --- |
|  | Alternative Hypothesis | z (by hand) | t (know equation only) |
| upper tailed\* | Ha: μ > μ0 (Δ0) | P(Z > zts) = 1 – P(Z ≤ zts) | P(T > ts) |
| lower tailed\* | Ha: μ < μ0 (Δ0) | P(Z < zts) | P(T < ts) |
| two-tailed | Ha: μ ≠ μ0 (Δ0) | 2(P(Z > | zts |)) = 2(1 – P(Z < | zts|)) = 2 P(Z < -|zts|) | 2(P(T > |ts|)) |

P value, or calculated probability, is the probability of finding the observed, or more extreme, results when the null hypothesis (H 0) of a study question is true

画出 p 的图 alpha beta

1. Define and interpret the power (Power = 1 – β) and state what factors affect the power and what the effects are.
2. Be able to calculate the power given the Ha, α and the true value of the mean, μa.

Significant： 成功reject, 不含0

Analysis of Variance (ANOVA): test differences between two or more means. Because inferences about means are made by analyzing variance.

ANOVA 前提：

a) Normality: QQ plot, histogram

b) Constant standard deviation: the ratio of standard deviations.

**ANOVA**

|  |  |  |  |
| --- | --- | --- | --- |
| Source | df | SS | MS |
| Factor A  (between) | k-1 |  |  |
| Error  (within) | n-k |  |  |
| Total | n-1 |  |  |

Test statistics F = , df1 = dfa, df2 = dfe

SST=SSA+SSE dft = dfa+dfe

P(F>Fdf1,df2) two-side

|  |  |
| --- | --- |
| CI diff sample size |  |
| same sample size |  |

**Constant variance**

**Bonferroni** (all comparisons, very conservative)

**Dunnett** (comparing to control only)

**Tukey** (All comparisons…this is the one we will do by hand)

To perform Tukey’s multiple comparison analysis indicating the factors that are different

a)

b)

c)

1. Be able to state the advantages for using a t distribution versus an F distribution and why an F distribution would be preferred.

t: compare mean, only 2

F: compare ratio of variance, so can be more than 2

1. Be able to explain why you have to use a special technique to perform multiple comparison analysis.
2. Be able to determine when to use Tukey’s multiple comparison analysis.
3. Be able to perform Tukey’s multiple comparison analysis indicating the factors that are different, presenting a visual display and presenting the final results in ‘layman’s’ terms. Note because of the time involved in the calculations, you might just be required to do a), b), c) below and determine the visual display from the resulting confidence intervals.

a)

b)

d) Example visual display:

|  |  |  |
| --- | --- | --- |
| 0 mg (control) | 20 mg | 40 mg |
| 57.60 | 69.28 | 75.70 |
|  |  |  |

scatterplot in terms of pattern, direction, strength, outliers, constant variance.

1. Be able to state the model for linear regression defining all of the terms and the assumptions for ε

Y = β0 + β1X + ε

least squares regression line:

Residual: ei = yi - ŷi

standard deviation about the least squares line:

coefficient of determination:

1. Be able to interpret R2 including what it doesn’t tell you.

a) linearity

b) outliers

c) good prediction

1. Be able to state the four assumptions for linear regression

a) Given the appropriate graphs, determine if the assumptions are met (you will need to determine which graphs are appropriate)

|  |  |  |  |
| --- | --- | --- | --- |
| Source | df | SS  (Sum of Squares) | MS  (Mean Square) |
| Regression | 1 |  |  |
| Error | n - 2 |  |  |
| Total | n - 1 |  |  |

a) SST = SYY

b) SSR = b1 SXY

c) SST = SSR + SSE

b) dft = dfr + dfe

perform hypothesis test for association (model utility test) with a test statistic of

confidence interval for β1:  ; df = dfe = n - 2

hypothesis tests for β1 with a test statistic of :

sample correlation:

1. Be able to interpret r

a) What happens when you switch X and Y

b) Sign of r

c) What is meant by uncorrelated (r = 0)

d) How r relates to the association of X and Y

e) Correlation doesn't provide information on form.

1. Be able to determine and explain when you cannot use linear regression
2. Be able to state if there is causality in linear regression.
3. Be able to calculate the confidence interval for the mean value at x = x\*.

a) The estimated value is the ŷx\*

c) df = n - 2

1. Be able to calculate the prediction interval at x = x\*.

a) The estimated value is the ŷx\*

c) df = n - 2

1. Be able to state the difference between the confidence interval for the mean response at x = x\* and the prediction interval for a particular value and when each would be used.