Compiled by Samantha Mukonjia (2020)

**BIOSTATISTICS - formulae**

Define the following terms – Probability, Experiment, outcomes, Sample Space, Event **1. Probability**

(i) Events

▪ Of two mutually exclusive events 🡪 ***p* (A**∪**B) = *p(A)* + *p(B) = 1***

▪ Of two independent events-- ***p*(A** ∩ **B) = *p(A)\*p(B)***

(ii) Probability distribution of variables

1. Normal distribution - (Continuous variables)-- First standardize variable to standard normal deviate (Z-score), **then interpret Z-Score**

�� =�� − ��

��

2. Binomial Distribution - (Binary Variables)

��(�� = ��) = (����)����(�� − ��)��−��

Where:

(����) =��!

��! (�� − ��)!

Where n (number of trials/observations) is large the distribution of the binomial variable and the proportion are approximately **normal when:**

���� ≥ ��

and ��(1 − ��) ≥ ��

The mean and variance of the binomial distribution can then be calculated as follows:

�� = ����

���� = ����(�� − ��)

��ℎ�� ���������������� ������ ������������ ������������������������ ������ ��ℎ���� ���� ��������������

3. Poisson Distribution- (Discrete/Rare occurrences)

�� (�� = ��) =������−��

��!

**2. Statistical Inference**

***Conclusions about a population are made based on findings from sample of the population.*** Measures of Statistical Inference:

∙ Confidence Interval (Commonly at 95% CI)

**CI of Means** (use Z-score) ***Use t distribution table when sample size < 20***

���� = ��̄± ��. ���� × ���� (��̄)

**1 |** P a g e

Compiled by Samantha Mukonjia (2020)

**Standard Error of mean =**����

√��

**CI of Proportions** (use Z-score) ***Use t distribution table when sample size < 20*** ���� = ��̄± ��. ���� × ���� (��̄)

**Standard Error of Proportion =** √��(��−��)

��

∙ Hypothesis Testing (P of 0.05 is usually used)

o State Hypothesis

����: �� ≠ ���� (Two sided Hypothesis)

����: �� > ����- (One sided Hypothesis)

����: �� < ����- (One Sided Hypothesis)

***Use z tables, but t distribution if n<20***

�� =��̄− ����

����(��)�� =��̄− ����

����(��)

�� =��̄− ����

����(��)

**Errors in Hypothesis testing**

| ���� | | |
| --- | --- | --- |
|  | True | False |
| **Accept** | �� − �� **(confidence interval)** | Type II error (��) |
| **Reject** | Type 1 error (��) | �� − ��**(power)** |

For a given sample size (n), lowering �� increases ��. Probability of a type II error decreases with increases in n.

**3. Comparing two means and proportions**

Is there a statistically significant difference? (Sample size>20)

**A. Means**

**1. Confidence Interval**

���� = (��̄��− ��̄��) ± ������× √(������

����+������

����)

**2. Significance testing – State hypothesis first.**

Interpret the Z score

�� =(��̄�� − ��̄��) √(������

����+������

����)

Is there a statistically significant difference? (Sample size<20) 🡪 Use t distribution (Additional Assumption is that there is commonality of variance)

**2 |** P a g e

Compiled by Samantha Mukonjia (2020)

Then

**1. Confidence Interval**

���� = (��̄�� − ��̄��) ± ��∗ × √ ������(������+������)

**Where**

������ =(���� − ��) ������(���� − ��)������

**2. Significance test**

(���� + ���� − ��)

**a. Unpaired t- test (Assumes commonality of variance)** �� =(��̄�� − ��̄��)

√ ������(��~~��~~��+��~~��~~��)

**b. Welch t-test (If the variances are different\_ Unequal Variances)**

**(i) First: Test for equality of variances using F-test F-test**

��(����,����) =������

������

���� ������( ���� = ���� − ��), ( ���� = ���� − ��)

**If the test is significant do the welch t-test**

**(ii) Second: Welch t-test (Resembles the Z test)**

�� =(��̄�� − ��̄��)

√(������

~~��~~��+������

~~��~~��)

**c. What if the data is paired?** Observations are not independent, however different pairs are independent

**(i) Paired t-test**

Steps

1. Calculate the differences between the observations on each pair (Distinguish positives and negatives)

2. Calculate the mean difference ��̄

3. State the hypothesis

4. Calculate standard deviation of the differences ����

5. From the above, calculate the **Standard Error of the Difference** ̄ = ����

����(��)

√��

�� =��̄

����( ��̄ )

6. Calculate t statistic

7. Interpret t value. Compare your value to the critical ����−�� distribution 🡪 this will give the p-value for the paired t test.

**3 |** P a g e

Compiled by Samantha Mukonjia (2020)

**B. Proportions**

**1. Confidence Interval**

���� = �� �������������������� ± ��. ���� × ���� (��������������������)

����(��������������������) = √[����(�� − ����)

����+����(�� − ����)

����]

**2. Statistical Significance**

�� =(���� − ����)

√��̄��̄(��~~��~~��+��~~��~~��)

**Where** ��̄ =����+����

**4. Association of two categorical variables 1. Chi-square test**

����+����

**Assumptions – Sample size>40 & smallest expected value** ≥ ��

**If n is between 20& 40 & smallest expected value is at least 5**

(i) Cross tabulate – Exposures and Outcomes

|  | Disease + | Disease - |  |
| --- | --- | --- | --- |
| Factor + | a | b | (a+b) |
| Factor - | c | d | (c+d) |
|  | (a+c) | (b+d) | n |

(ii) State the hypothesis

(iii) Chi-square formulae

���� = ∑(�� − ��)��

��

�� =������ ���������� �� ������������ ����������

�������������� ����������

(iv) Establish degree of freedom = (r-1)x(c-1)

(v) Check the Chi-square matching that degree of freedom at 95% ��

confidence interval ����.����

(vi) Compare the two: if less than critical value 🡪 accept the null hypothesis **2. If the above assumptions don’t apply – use** Fisher’s exact test.(Only for 2 by 2 tables) **Steps**

(i) Cross Tabulate exposures and outcomes

| **Exposure** | Disease + | Disease - | (**Outcomes**) |
| --- | --- | --- | --- |
| Factor + | a | b | (a+b) |
| Factor - | c | d | (c+d) |
|  | (a+c) | (b+d) | n |

(ii) State the hypothesis

**4 |** P a g e

Compiled by Samantha Mukonjia (2020)

**(iii) Do Fisher’s exact test**

(�� + ��

��) + (�� + ��

�� =

(~~��~~

�� + ��)

��)

**5 |** P a g e