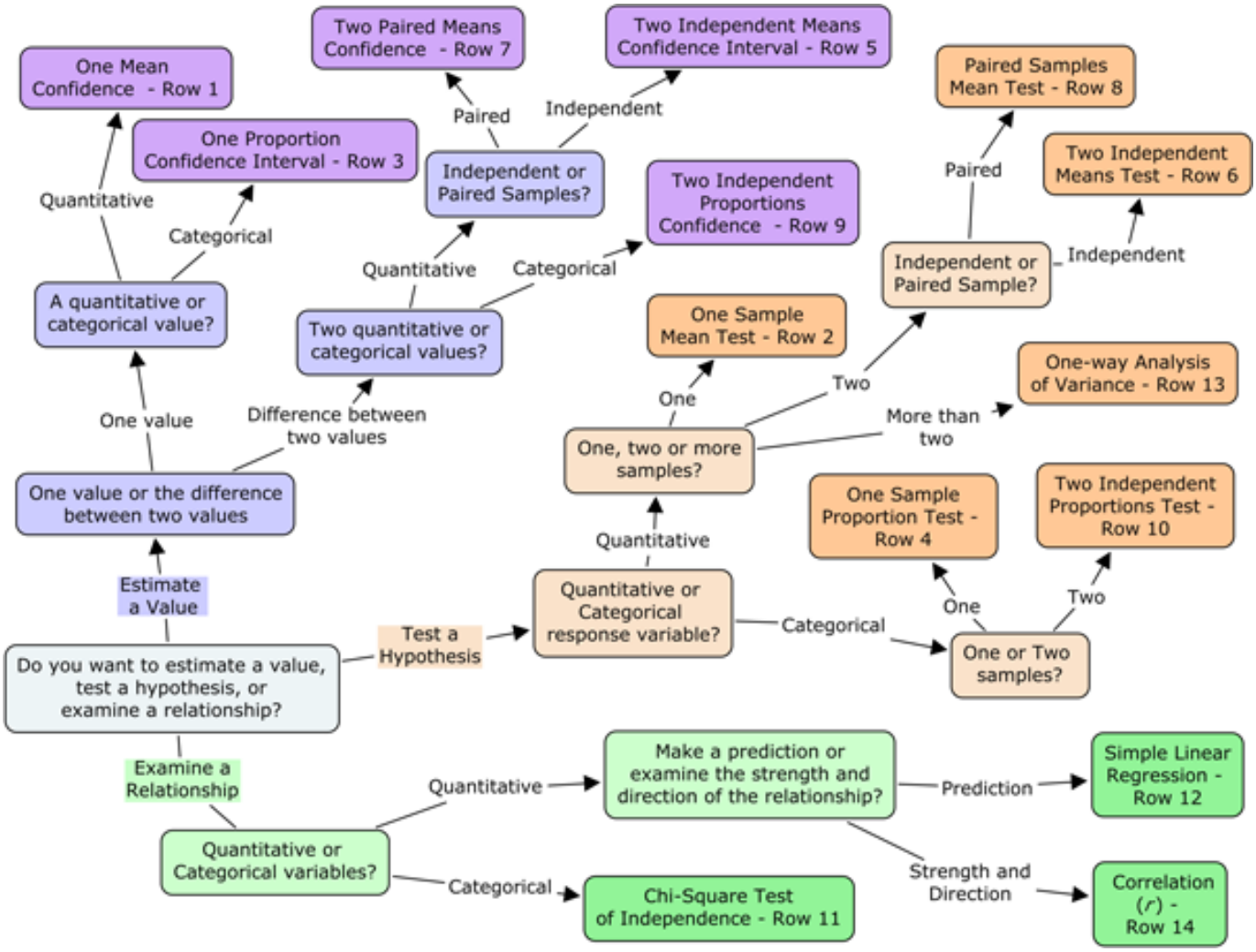
greatlearning

Summary Diagram and Information regarding Application of Statistical Procedures[i]

The rows correspond to the table below



greatlearning

Table of Recommended Procedures

1. Estimate a population mean μ, using a t-statistic.
2. Test of a hypothesis about a population mean μ, using a t-statistic.
3. Estimate a population proportion π using a z-statistic.
4. Test about a population proportion π using a z-statistic.
5. Estimate the differences between two population means μ_1 and μ_2 for independent samples using a 2-sample t-statistic. You also need to specify whether to use **pooled variance** or **separate variance**.
6. Test to compare two population means μ_1 and μ_2 for independent samples using a 2-sample t-statistic. You also need to specify whether to use **pooled variance** or **separate variance**.
7. Estimate the difference of two means in a paired comparison study using a t-statistic.
8. Test to compare two means in a paired comparison study using a t-statistic.
9. Estimate of difference of two populations proportions π_1 and π_2 using a z-statistic.
10. Test to compare two population proportions π_1 and π_2 using a z-statistic.
11. Test the dependence of two categorical variables using a Chi-square test of independence.
12. Finding the best fitting line to a set of data with a quantitative explanatory variable X and a quantitative response variable Y and examining the slope of the regression line.
13. Test to compare several population means.
14. Test whether two population variances are different using the homogeneity of variances test (F test if normality is assumed or Levene's test if normality is not assumed).

Items 12 & 13 are to be covered in AS.

Application of the Concepts

1. A survey of National Federation of Independence Business (NFIB) indicates that small businesses intended to increase their hiring as well as their capital expenditures during 1986 as compared with 1985. Suppose that, as part of a follow-up survey by NFIB, 20 small businesses, randomly chosen from the NFIB's list of 2,100 companies, show an average hiring from 1985 equal to 3.2 new employees per firm and a standard deviation of 1.5 hires. A random sample of 30 small businesses taken at the end of 1986 shows an average of 5.1 new hires and a standard deviation of 2.3 hires. At α=0.01 level of significance, can you conclude that average hiring by all small businesses in 1986 increased as compared with 1985?

Test of a hypothesis about two independent means μ1–μ2, using a t-statistic after testing whether the population variances are equal.

Test for equality of variances: H0: $\sigma_1^2 = \sigma_2^2$ against the alternative Ha: $\sigma_1^2 \neq \sigma_2^2$

$S_1^2 / S_2^2 = 0.4253$

```
> qf(0.05,19,29)
[1] 0.4814141
> qf(0.995,19,29)
[1] 2.884736
```

H0 rejected. Equality of population variances cannot be assumed.

H0: $\mu_1 = \mu_2$ against the alternative Ha: $\mu_1 < \mu_2$

T = -3.53

DF = 47.99 ≈ 48

```
> qt(0.01, 48)
[1] -2.406581
```

H0 rejected. Equality of population means cannot be assumed. Hiring shows significant increase in 1986 from 1985.

2. It is known that the average stay of tourists in Hong Kong hotels has been 3.4 nights. A tourism industry analyst wanted to test whether recent changes in the nature of tourism to Hong Kong have changed from this past average. The analyst obtained the following random sample of the number of nights spent by tourists in Hong Kong hotels: 5, 4, 3, 2, 1, 1, 5, 7, 8, 4, 3, 3, 2, 5, 7, 1, 3, 1, 1, 5, 3, 4, 2, 2, 2, 6, 1, 7. Conduct the test using the 0.05 level of significance.

Test of a hypothesis about population mean μ, using a t-statistic.

H0: μ= 3.4 against the alternative Ha: μ ≠ 3.4

```
> t.test(ngt, mu=3.4, alternative="two.sided")

  One Sample t-test

data:  ngt
t = 0.24792, df = 27, p-value = 0.8061
alternative hypothesis: true mean is not equal to 3.4
95 percent confidence interval:
 2.672376 4.327624
sample estimates:
mean of x
3.5
```

At 5% level of significance the null hypothesis is not rejected.

3. There are 155 banks involved in certain international transactions. A federal agency claims that at least 35% of these banks have total assets of over $10 billion (In U.S. dollars). An independent agency wants to test this claim. It gets a random sample of 50 out of the 155 banks and finds that 15 of them have total assets of over $10 billion. Can the claim be rejected?

Test about a population proportion π using a z-statistic.

H0: $\pi \geq 0.35$ against the alternative Ha: $\pi < 0.35$

```
> prop.test(15, 50, 0.35, alternative="less")

        1-sample proportions test with continuity correction

data:  15 out of 50, null probability 0.35
X-squared = 0.35165, df = 1, p-value = 0.2766
alternative hypothesis: true p is less than 0.35
95 percent confidence interval:
 0.0000000 0.4249925
sample estimates:
  p
0.3

> prop.test(15, 50, 0.35, alternative="less", correct=F)

        1-sample proportions test without continuity
        correction
```

```
data:  15 out of 50, null probability 0.35
X-squared = 0.54945, df = 1, p-value = 0.2293
alternative hypothesis: true p is less than 0.35
95 percent confidence interval:
 0.0000000 0.4145996
sample estimates:
  p
0.3
```

The claim cannot be rejected at 5% level of significance.

4. General Motors Corporation hopes to reduce anticipated production costs of its Saturn Model by instituting an assembly schedule that will reduce average production time to about 40 hours per car. In a test run of the new assembly line, 40 cars are built at a sample average time per car of 46.5 hours and a sample standard deviation of 8.0 hours. A test run of 38 cars using the old assembly schedule results in a sample mean of 51.2 hours and a sample deviation of 9.5 hours. Is there proof that the new assembly schedule reduces the average production time per car? What is the *p*-value? Explain.

Test to compare two population means μ1 and μ2 for independent samples using a 2-sample t-statistic. You also need to specify whether to use pooled t-test or nonpooled t-test.

5. A telephone company wants to estimate the average length of long-distance calls during weekends. A random sample of 50 calls gives a mean of 14.5 min and standard deviation $s = 5.6$ min. Give a 95% confidence interval for the average length of a long-distance phone call during weekends.

Estimate a population mean μ, using a t-statistic.

95% CI for μ: $\bar{x} \pm t(49, 0.975) \, {}^{s}/_{\sqrt{n}}$

```
> qt(0.975, 49)
[1] 2.009575
> 14.5-2.009575*5.6/sqrt(50)
[1] 12.9085
> 14.5+2.009575*5.6/sqrt(50)
[1] 16.0915
```

Interval: [12.91, 16.09]

6. Several companies have been developing electronic guidance systems for cars. Motorola and Germany's Blaupunkt are two firms in the forefront of such research. Out of 120 trials of the

Motorola model, 101 were successful; and out of 200 tests of the Blaupunkt model, 110 were successful. Is there evidence to conclude that the Motorola electronic guidance system is superior to the German competitor?

Test about two population proportion π1 and π2 using a z-statistic

H0: $\pi_1 = \pi_2$ against the alternative Ha: $\pi_1 > \pi_2$

```
> prop.test(c(101, 110), c(120, 200), alternative="greater")

        2-sample test for equality of proportions with
        continuity correction

data:  c(101, 110) out of c(120, 200)
X-squared = 27.123, df = 1, p-value = 9.544e-08
alternative hypothesis: greater
95 percent confidence interval:
 0.2052961 1.0000000
sample estimates:
   prop 1    prop 2
0.8416667 0.5500000
```

Null hypothesis is rejected. Evidence points to Motorola guidance system to be better than German system.

7. An important measure of the risk associated with a stock is the standard deviation, or variance, of the stock's price movements. A financial analyst wants to test the one-tailed hypothesis that stock A has a greater risk (larger variance of price) than stock B. A random sample of 25 daily prices of stock A gives s^2_A=6.52, and a random sample of 22 daily prices of stock B gives a sample variance of s^2_B=3.47. Carry out the test at α=0.01.

Test whether two population variances are different using the homogeneity of variances

8. A company is interested in offering its employees one of two employee benefit packages. A random sample of the company's employees is collected, and each person in the sample is asked to rate each of the two packages on an overall preference scale of 0 to 100. The order of presentation of each of the two plans is randomly selected for each person in the sample. The paired data are:

Program A: 45 67 63 59 77 69 45 39 52 58 70 46 60 65 59 80
Program B: 56 70 60 45 85 79 50 46 50 60 82 40 65 55 81 68

Test to compare two means in a paired comparison study using a t-statistic.

H0: $\mu_1 = \mu_2$ against the alternative Ha: $\mu_1 \neq \mu_2$

```
        Paired t-test

data:  X and Y
t = -0.97511, df = 15, p-value = 0.345
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -7.566422  2.816422
sample estimates:
mean of the differences
                -2.375
```

H0 cannot be rejected. The two population means cannot assumed to be different.

9. The maker of portable exercise equipment, designed for the health-conscious people who travel too frequently to use a regular athletic club, wants to estimate the proportion of traveling business people who may be interested in the product. A random sample of 120 traveling business people indicates that 28 of them may be interested in purchasing the portable fitness equipment. Give a 95% confidence interval for the proportion of all traveling business people who may be interested in the product.

Estimate a population proportion π using a z-statistic.

```
> prop.test(28, 120)

        1-sample proportions test with continuity correction

data:  28 out of 120, null probability 0.5
X-squared = 33.075, df = 1, p-value = 8.867e-09
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.1630688 0.3210327
sample estimates:
        p
0.2333333
```

Required CI: [0.16, 0.32]

10. When new paperback novels are promoted at bookstores, a display is often arranged with copies of the same book with differently colored covers. A publishing house wanted to find out whether there is a dependence between the place where the book is sold and the color of its cover. For one of its latest novels, the publisher sent displays and a supply of copies of the novels to large bookstores in five major cities. The resulting sales of the novel for each city-color combination are as follows. Numbers are in thousands of copies sold over a three-month period.

	Color				
City	**Red**	**Blue**	**Green**	**Yellow**	**Total**
New York	21	27	40	15	103
Washington	14	18	28	8	68
Boston	11	13	21	7	52
Chicago	3	33	30	9	75
Los Angeles	30	11	34	10	84
Total	79	102	153	49	383

Assume that the data are random samples for each particular color-city combination and that the inference may apply to all novels. Conduct the overall test for independence of color and location.

Test the dependence of two categorical variables using a Chi-square test of independence.

H_0: City and Color are independent against alternative Ha: City and Color are not independent

```
> chisq.test(Bk)

    Pearson's Chi-squared test

data:  Bk
X-squared = 33.958, df = 12, p-value = 0.0006849
```

Conclusion: City and color are not independent.

11. Certain eggs are stated to have reduced cholesterol content, with an average of only 2.5% cholesterol. A concerned health group wants to test whether the claim is true. The group believes that more cholesterol may be found, on the average, in the eggs. A random sample of 100 eggs reveals a sample average content of 5.2% cholesterol, and a sample standard deviation of 2.8%. Does the health group have cause for action?

Test of a hypothesis about a population mean μ, using a t-statistic.

12. Two 12-meter boats, the K boat and the L boat, are tested as possible contenders in the America's Cup races. The following data represent the time, in minutes, to complete a particular tack in independent random trials of the two boats.
    K boat: 12.0, 13.1, 11.8, 12.6, 14.0, 11.8, 12.7, 13.5, 12.4, 12.2, 11.6, 12.9
    L boat: 11.8, 12.1, 12.0, 11.6, 11.8, 12.0, 11.9, 12.6, 11.4, 12.0, 12.2, 11.7
    Test the null hypothesis that the two boats perform equally well. Is one boat faster, on the average, than the other? Assume equal population variances.

Test to compare two population means μ_1 and μ_2 for independent samples using a 2-sample t-statistic. You also need to specify whether to use pooled t-test or non-pooled t-test.

13. A study undertaken by Montgomery Securities to access average labor and materials costs incurred by Chrysler and General Motors in building a typical four-door, intermediate-sized car. The reported average cost for Chrysler was $9500, and for GM it was $9780. Suppose that these data are based on random samples of 25 cars for each company, and suppose that both standard deviations are equal to $1500. Test the hypothesis that the average GM car of this type is more expensive to build than the average Chrysler car of the same type. Use $\alpha=0.01\alpha=0.01$. Assume equal population variances.

Test to compare two population means μ_1 and μ_2 for independent samples using a 2-sample t-statistic. You also need to specify whether to use pooled t-test or non-pooled t-test.

14. Recent studies indicates that in order to be globally competitive, firms must form global strategic partnerships. An investment banker wants to test whether the return on investment for international ventures is different from return on investment for similar domestic ventures. A sample of 12 firms that recently entered into ventures with foreign companies is available. For each firm, the return on investment for both the international venture (I), and similar domestic venture (D) is given:
    D(%):10 12 14 12 12 17 9 15 8.5 11 7 15
    I(%) : 11 14 15 11 12.5 16 10 13 10.5 17 9 19

Assuming that these firms represent a random sample from the population of all firms involved in global strategic partnerships, can the investment banker conclude that there are differences between average returns on domestic ventures and average returns on international ventures? Explain.

Test to compare two means in a paired comparison study using a t-statistic.

[i] Source: Penn State World Campus Stat 500 Course Notes