Elementary Statistics

Formulas

Elementary Statistics, Exam 1

Introduction to Statistics, Exploring Data with Tables and Graphs, Describing, Exploring, and Comparing Data

$Range = High - Low$, $W = \frac{Range}{\#\ of\ classes}$ →**Always round up if a remainder.**

The class width: W = UL – LL + 1 = UB – LB = LL of 2nd – LL of 1st

The *class midpoint* : $Midpoint = \frac{UL + LL}{2} = \frac{UB + LB}{2}$

Relative Frequency (RF) = *f* / *n*

Mean: $\bar{x} = \frac{\sum x}{n}$, $\mu = \frac{\sum X}{N}$

Median: **The middle value of ranked data**

Mode: **The value(s) that occur(s) with the greatest frequency.**

Midrange: $Mr = \frac{Min + Max}{2}$

Weighted Average: $\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n} = \frac{\sum wx}{\sum w}$

5-Number Summary:

Minimum, Q_1, Q_2 , Q_3, Maximum

$IQR = Q_3 - Q_1$

Outliers = Values < $Q_1 - 1.5IQR$

OR

Outliers = Values > $Q_3 + 1.5IQR$

TI Calculator:

How to enter data:

1. Stat
2. Edit
3. ClrList L_1
4. Or Highlight & Clear
5. Type in your data in L1, ..

Mean, SD, 5-number summary

1. Stat
2. Calc
3. Select 1 for 1 variable
4. Type: L1 (second 1)
5. Scroll down for 5-number summary

$Percentile = \frac{(\#\text{ of values below } X)}{\text{total \# of values}} \cdot 100\%$

L locator :

$L = \frac{k}{100} \cdot n = \frac{P}{100} \cdot n$

The linear correlation coefficient: $-1 \le r \le 1$

Variance

$$\sigma^2 = \frac{\sum (X-\mu)^2}{N}$$

$$s^2 = \frac{n\sum X^2 - \left(\sum X\right)^2}{n(n-1)}$$

$$s^2 = \frac{\sum (X-\bar{X})^2}{n-1}$$

Standard Deviation

$$\sigma = \sqrt{\frac{\sum (X-\mu)^2}{N}}$$

$$s = \sqrt{\frac{\sum (X-\bar{X})^2}{n-1}}$$

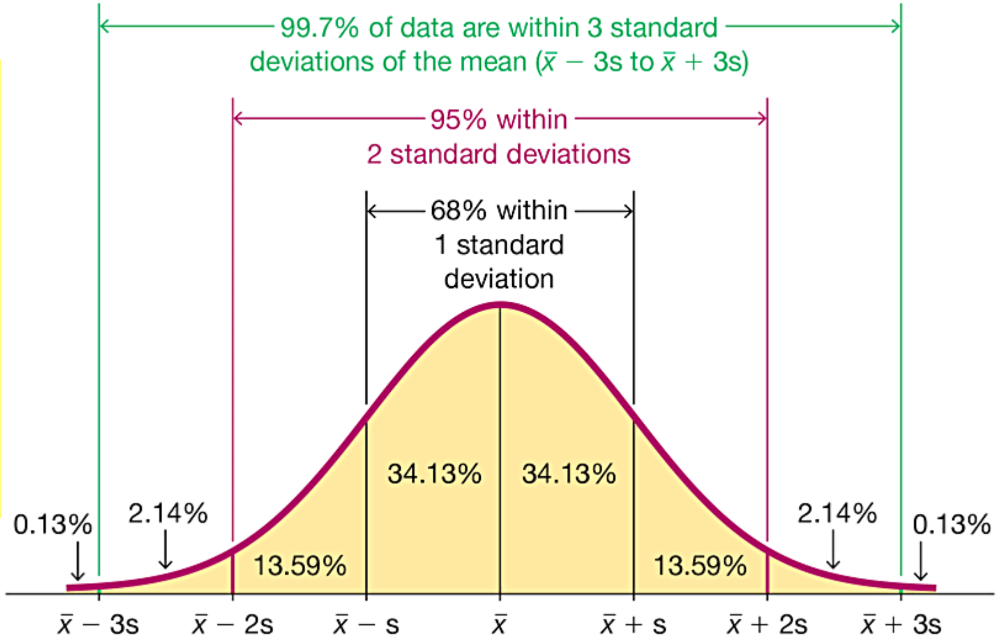
$$s = \sqrt{\frac{n\sum X^2 - \left(\sum X\right)^2}{n(n-1)}}$$

Z – Scores:

$$z = \frac{x-\mu}{\sigma}$$

$$z = \frac{x-\bar{x}}{s}$$

Empirical Rule (Normal)



Grouped Data: *m* is the Midpoint of a class

$$\mu = \frac{\sum mf}{N}, \sigma^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{N}}{N} \Rightarrow \sigma = \sqrt{\frac{\sum m^2 f}{N} - \mu^2}$$

$$OR: \bar{x} = \frac{\sum f \cdot x_m}{n}$$

Chebyshev's Theorem: $1 - 1/k^2$

Coefficient of Variation

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Range = Max – Min

Range Rule of Thumb :

$s \approx \frac{Range}{4}$ & $\bar{x} \pm 2s$ ***OR*** $\mu \pm 2\sigma$

Unusual Values (Significantly low or high):

z **scores** ≤ -2.00 **or**

z **scores** ≥ 2.00

Elementary Statistics, Exam 2

Probability
Discrete Probability Distribution

$0 \le P(A) \le 1, \sum P_i = 1$

Impossible Set: $P(A) = 0$

Sure (Certain) Set: $P(E) = 1$

Complementary Events: $P(\bar{A}) = 1 - P(A)$

The actual odds against event A: $O(\bar{A}) = \frac{P(\bar{A})}{P(A)}$,

The actual odds in favor of event A: $O(A) = \frac{P(A)}{P(\bar{A})}$

$\text{Payoff odds against event } A = \frac{\text{net profit}}{\text{amount bet}}.$

$$n(s) = (\#\ of\ outcomes)^{\#\ of\ stages}$$

$$P(E) = \frac{n(E)}{n(S)}$$

$$or P(A) = \frac{s}{n}$$

Conditional probability

Prob of A Given B: $P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) \neq 0$

Prob of B Given A: $P(B|A) = \frac{P(A \cap B)}{P(A)}; P(A) \neq 0$

A & B are independent:

$P(A \cap B) = P(A) \cdot P(B)$

A & B are Dependent:

$P(A \text{ and } B) = P(A) \cdot P(B \mid A)$

$= P(B) \cdot P(A \mid B)$

Addition Rule (Union): $P(A \cup B) = P(A) + P(B) - P(A \text{ and } B)$

Disjoint (or mutually exclusive): $P(A \cap B) = 0$

$P($**at least one** occurrence of event $A) = 1 - P($**no** occurrences of event $A)$

$$\text{Factorial: } n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1, 0! = 1$$

$$\text{Permutation: } {}_nP_r = \frac{n!}{(n-r)!}$$

$$\text{Combination: } (n,r) = C_r^n = {}_nC_r = \frac{n!}{(n-r)!r!}$$

Permutations Rule (When Some Items Are Identical to Others)

n_1 are alike, n_2 are alike, . . . , and n_k are alike.

$$\frac{n!}{n_1!\cdot n_2!\cdots\cdot n_k!}$$

Parameters of a Probability Distribution:

Mean (Expected Value), Variance & Standard Deviation of D.R.V x:

$$\text{Mean: } \mu = E(x) = \sum x\cdot P(x)$$

$$\text{Variance: } \sigma^2 = \sum (x-\mu)^2 p(x) = \sum\left[x^2\cdot P(x)\right] - \mu^2$$

$$\text{SD: } \sigma = \sqrt{\sum (x-\mu)^2 p(x)} = \sqrt{\sum x^2 p(x) - \mu^2}$$

TI Calculator:
Factorial

1. Enter the value of n
2. Press Math
3. Select PRB
4. Select ! & Enter

TI Calculator:
Permutation / Combination

1. Enter the value of n
2. Press Math
3. Select PRB
4. Select nP_r, or nC_r
5. Enter the value of r & Enter

Binomial Probability Distributions

$$p(x) = \frac{n!}{(n-x)!x!}p^x q^{n-x}$$

$$\mu = np$$

$$\sigma = \sqrt{npq} = \sqrt{np(1-p)}$$

TI Calculator:
Binomial Distribution

1. 2nd + VARS
2. binompdf(
3. Enter: n, p, x
4. Enter
5. If you enter n, p only
6. Gives all probabilities from 0 to n
7. If using Binomcdf(
8. Gives sum of the probabilities from 0 to x.

Elementary Statistics, Exam 3

Normal Probability Distribution

Estimating Parameters and Determining Sample Sizes

Normal Distribution

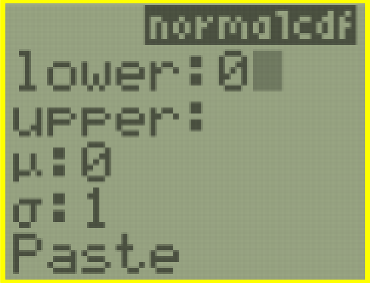
SND: 1) Bell-shaped, 2) $\mu = 0$, 3) $\sigma = 1$

Normal to SND: $z = \frac{x-\mu}{\sigma}$

TI Calculator:

Normal Distribution Area

1. **2^{nd} + VARS**
2. **normalcdf(**
3. **4 entries required**
4. **Left bound, Right bound, value of the Mean, Standard deviation**
5. **Enter**
6. **For $-\infty$, $use\ -1000$**
7. **For ∞, use 1000**



TI Calculator:

Normal Distribution: find the Z-score

1. **2^{nd} + VARS**
2. **invNorm(**
3. **3 entries required**
4. **Left Area, value of the Mean, Standard deviation**
5. **Enter**

The Central Theorem:

1. The distribution of sample $\bar{x}$ will, as the sample size increases, approach a normal distribution.
2. $\mu_{\bar{x}} = \mu$.
3. $\sigma_{\bar{x}} = \sigma/\sqrt{n}$
4. $z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Normal as Approximation to Binomial Requirements

1. The sample is a simple random sample with n independent trials of a binomial experiment with the probability of success is p. $p \to q = 1 - p$
2. $np \geq 5$ and $nq \geq 5$.

Normal Approximation: $\mu = np$ & $\sigma = \sqrt{npq}$

Continuity Correction:

Adjust the discrete whole number x by using a **continuity correction Factor: 0.5**

The **margin of error (maximum error of the estimate): *E***

Sample proportion: $\hat{p} = \frac{x}{n}$ (read p “hat”)

$\hat{q} = 1 - \hat{p}$

The population proportion $p = \frac{X}{N}$

Point estimate of *p*: $\hat{p} = \frac{UCL+LCL}{2}$, UCL: Upper Confidence Limit

Margin of error: $E = \frac{UCL-LCL}{2}$, LCL: Lower Confidence Limit

Confidence Interval for Estimating a Population Proportion *p*

$$\hat{p} \pm E$$

$$\hat{p} - E < p < \hat{p} + E$$

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

TI Calculator:
Confidence Interval: proportion

1. Stat
2. Tests
3. 1-prop ZINT
4. Enter: ***x, n*** **& CL**

Determining the Sample Size:

$$n = \frac{(z_{\alpha/2})^2\,\hat{p}\hat{q}}{E^2}$$

When no estimate of $\hat{p}$ is known: $\hat{p} = \hat{q} = 0.5 \rightarrow$

$$n = \frac{(z_{\alpha/2})^2\,0.25}{E^2}$$

Point estimate of μ: $\bar{x} = \frac{UCL+LCL}{2}$, UCL: Upper Confidence Limit

Margin of error: $E = \frac{UCL-LCL}{2}$, LCL: Lower Confidence Limit

TI Calculator:
Z - interval
1. Stat
2. Tests
3. Z - Interval
4. Enter Data or Stats ($\bar{x}$, s & CL)

Confidence Interval for Estimating a Population Mean with σ Known

$$E = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$CI: \bar{x} \pm E \rightarrow$ $\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

Determine the sample size n for a population mean μ: $n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$

Confidence Interval for Estimating a Population Mean with σ unknown

$$E = t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$\bar{x} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{x} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

TI Calculator:
T- Distribution: find the t-score
1. 2nd + VARS
2. invT(
3. 2 entries (Left Area,df)
4. Enter

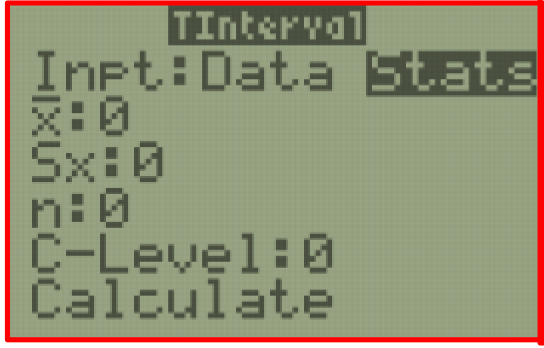
Confidence Interval for Estimating a Population Standard Deviation or Variance

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}},\ \text{d.f.} = n-1$$

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\text{right}}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{\text{left}}}}$$

TI Calculator:
How to enter data:
1. Stat
2. Edit
3. ClrList L_1
4. Or Highlight & Clear
5. Type in your data in L1, ..

TI Calculator:
How to enter data:
Mean, SD, 5-number summary
1. Stat
2. Calc
3. Select 1 for 1 variable
4. Type: L1 (second 1)
5. Scroll down for 5-number summary



TI Calculator:
T- interval
1. Stat
2. Tests
3. T - Interval
4. Enter Data or Stats ($\bar{x}$, s & CL)
5. Enter

Elementary Statistics, Exam 4

Hypothesis Testing
Inferences from Two Samples

$$\hat{p} = \frac{x}{n} \rightarrow x = n\hat{p}, \mu = np, \sigma = \sqrt{npq}$$

Parameter	Sampling Distribution	Requirements	*Test Statistics*
Proportion: p	Normal (Z)	$np \geq 5, \& \ nq \geq 5$	$Z = \frac{\hat{p} - p}{\sqrt{pq/n}}$
Mean: μ	t	σ not known & Normally Distributed Population or $n > 30$	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$
Mean: μ	Normal (Z)	σ known & Normally Distributed Population or $n > 30$	$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$
Standard Deviation: σ Or Variance: σ^2	χ^2	Normally Distributed Population	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

Step 1: H_0 , H_1, claim & Tails
Step 2: TS Calculate (**TS**)
Step 3: CV using α
Step 4: Make the **decision** to
a. Reject or not H_0
b. The claim is true or false
c. Restate this decision: There is / is not sufficient evidence to support the claim that…

TI Calculator:
Mean: T – Test
1. Stat
2. Tests
3. T – Test
4. Enter Data or Stats (p, x, n)
5. Choose RTT, LTT, or 2TT
6. Calculate

TI Calculator:
Mean: Z – Test
1. Stat
2. Tests
3. Z – Test
4. Enter Data or Stats (p, x, n)
5. Choose RTT, LTT, or 2TT
6. Calculate

TI Calculator:
1 - Proportion Z - test
1. Stat
2. Tests
3. 1 – PropZTest
4. Enter Data or Stats (p, x, n)
5. Choose RTT, LTT, or 2TT

Inferences about 2 Proportions

$$\hat{p}_1 = \frac{x_1}{n_1},\ \hat{q}_1 = 1 - \hat{p}_1$$

$$\hat{p}_2 = \frac{x_2}{n_2},\ \hat{q}_2 = 1 - \hat{p}_2$$

Pooled Sample Proportion :

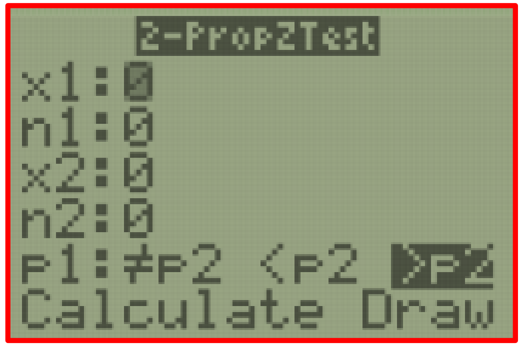
$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2},\ \bar{q} = 1 - \bar{p}$$

$$TS: z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{Or} \quad \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\,\bar{q}}{n_1} + \frac{\bar{p}\,\bar{q}}{n_2}}}$$

TI Calculator:

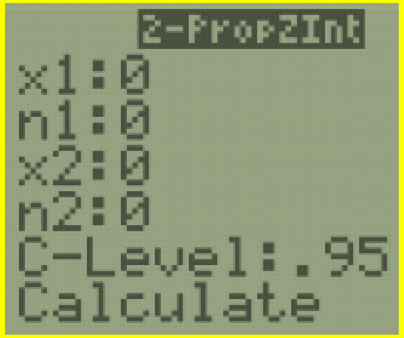
2 - Proportion Z - test

1. **Stat**
2. **Tests**
3. **2 – PropZTest**
4. **Enter Data or Stats** n_1, n_2, x_1, x_2
5. **Choose RTT, LTT, or 2TT**



Confidence Interval Estimate of $p_1 - p_2$

$$(\hat{p}_1 - \hat{p}_2) \pm E$$



TI Calculator:

Confidence Interval: 2 proportion

1. **Stat**
2. **Tests**
3. **2-prop ZINT**
4. **Enter:** n_1, n_2, x_1, x_2 **& CL**

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

The *P*-value method and the critical value method are equivalent, but the confidence interval method is **not** equivalent to the *P*-value method or the critical value method.

Inferences about 2 Means: Independent Samples

1. The two samples are **independent.** 2. Both samples are **simple random samples**. 3. Either or both of these conditions are satisfied: The two sample sizes are both **large** (with $n_1 > 30$ and $n_2 > 30$) or both samples come from populations having normal distributions.

σ_1 and σ_2 are known: Use the ***z* test** for comparing two means from independent populations

$$TS: z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{Or } z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Confidence Interval:

$$(\bar{x}_1 - \bar{x}_2) \pm E$$

$$E = z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

σ_1 and σ_2 are unknown: Use the ***t* test** for comparing two means from independent populations

Unequal Variances: $\sigma_1 \neq \sigma_2$

df = smaller of $n_1 - 1$ or $n_2 - 1$

$$TS: t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$E = t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Equal Variances :

$\sigma_1 = \sigma_2$

Pool the Sample Variances

df = $n_1 - 1 + n_2 - 1$

$$TS: t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

$$s_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

$$E = t_{\alpha/2}\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

TI Calculator:

2 - Sample Z - test

1. **Stat**
2. **Tests**
3. **2 – SampZTest**
4. **Enter Data or Stats** $\sigma_1, \sigma_2, \bar{x}_1, n_1, \bar{x}_2$ $n_1, n_2,$
5. **Choose RTT, LTT, or 2TT**
6. **Calculate**

TI Calculator:

2 - Sample Z - Interval

1. **Stat**
2. **Tests**
3. **2 – SampZInt**
4. **Enter Data or Stats** $\sigma_1, \sigma_2, \bar{x}_1, n_1, \bar{x}_2$ $n_1, n_2,$
5. **Choose RTT, LTT, or 2TT**
6. **Calculate**

TI Calculator:

2 - Sample T - test

1. **Stat**
2. **Tests**
3. **2 – SampTTest**
4. **Enter Data or Stats** $\bar{x}_1, s_1, n_1, \bar{x}_2$ $n_1, s_2,$
5. **Choose RTT, LTT, or 2TT**
6. **Pooled: No / Yes**
7. **Calculate**

TI Calculator:

2 - Sample T - Interval

1. **Stat**
2. **Tests**
3. **2 – SampTInt**
4. **Enter Data or Stats** $\bar{x}_1, s_1, n_1, \bar{x}_2$ $n_1, s_2,$
5. **Choose RTT, LTT, or 2TT**
6. **Pooled: No / Yes**
7. **Calculate**

Inferences about 2 Means: Dependent Samples (Matched Pairs)

The data consist of **matched pairs** (matched according to some relationship, such as before/after measurements from the same subjects)

- d = individual difference between the two values in a single matched pair
- μ_d = mean value of the differences d for the **population** of all matched pairs of data
- $\bar{d}$ = mean value of the differences d for the paired sample data
- s_d = standard deviation of the differences d for the paired **sample** data
- n = number of **pairs** of sample data

$$\text{d.f.} = n-1$$

Use either d or D:

$$\bar{D} = \frac{\sum D}{n} \quad \text{Or} \quad \bar{d} = \frac{\sum d}{n}$$

$$s_d = \sqrt{\frac{\sum (d-\bar{d})^2}{n-1}} = \sqrt{\frac{n\sum d^2 - (\sum d)^2}{n(n-1)}}$$

TI Calculator:
How to enter data:

1. **Stat**
2. **Edit**
3. **ClrList L_1 & L_2**
4. **Type in your data in L_1 & L_2**
5. **$L_1 - L_2$**
6. **Store in L_3**
7. **Enter**

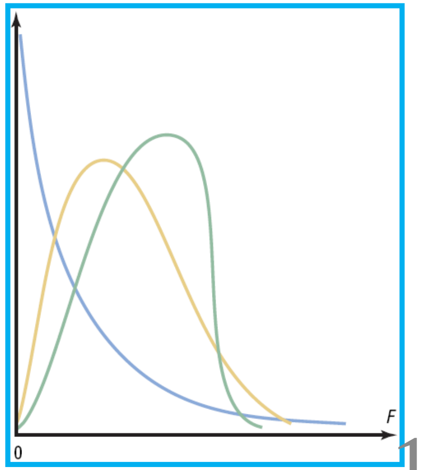
Mean, SD, 5-number summary

1. **Stat**
2. **Calc**
3. **Select 1 for 1 variable**
4. **Type: L3 (second 3)**
5. **Calculate**

Inferences about 2 variances or standard deviations:

The F test should not be confused with the chi-square test, which compares a single sample variance to a specific population variance.
The larger of the two variances is placed in the numerator regardless of the subscripts.
The F test has two terms for the degrees of freedom: that of the numerator, $n_1 - 1$, and that of the denominator, $n_2 - 1$, where n_1 is the sample size from which the larger variance was obtained.

Test Statistic: $F = \frac{s_1^2}{s_2^2}$



TI Calculator:
T- interval

1. **Tests**
2. **T - Interval**
3. **Data**
4. **Enter μ_0 =0, List: L_3, Freq:1**
5. **Calculate**

Confidence Interval:
$\bar{d} \pm E$ or $\bar{D} \pm E \rightarrow$

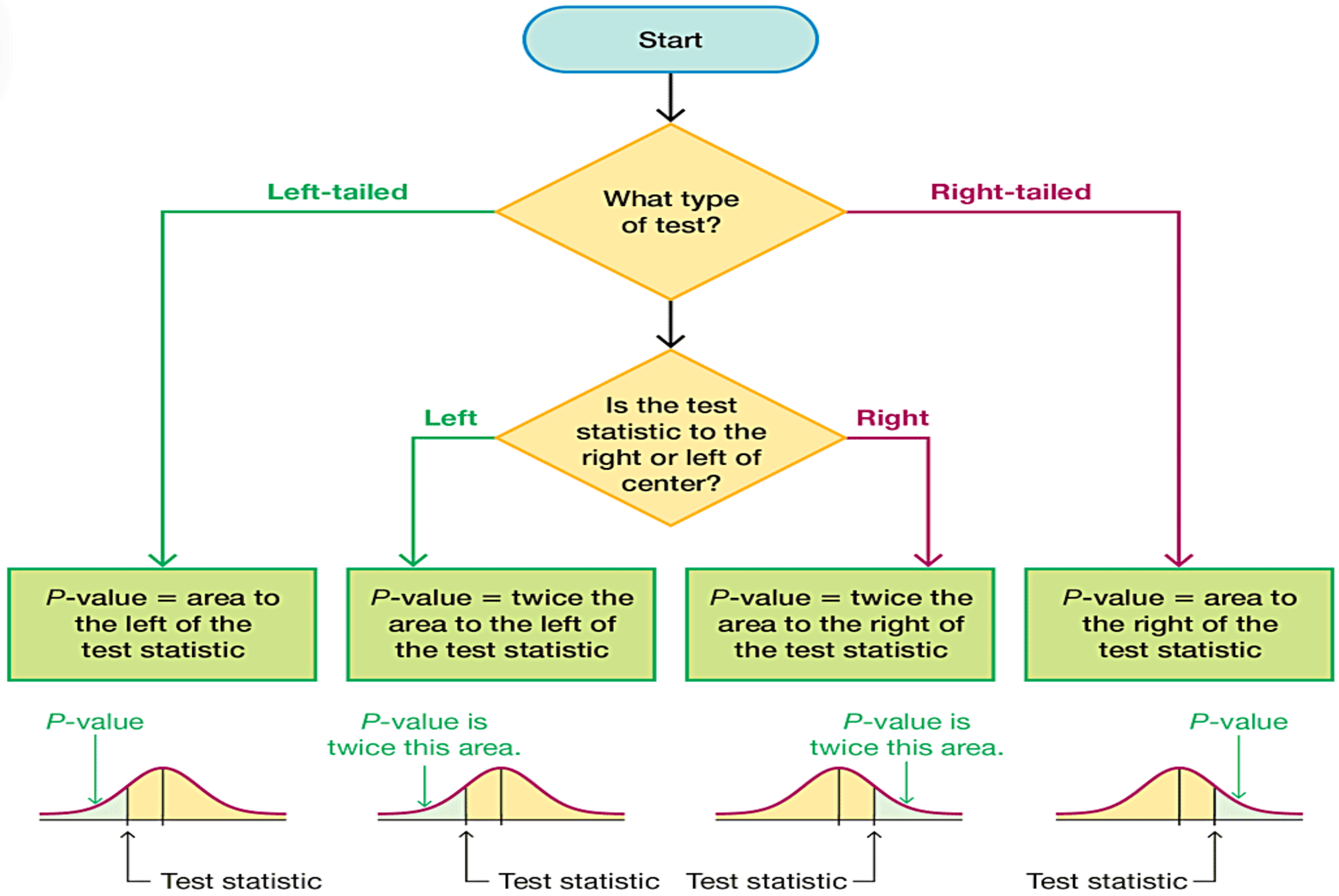
$$E = t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

$$TS: t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$$

TI Calculator:
Matched pair: T – Test

1. **Tests**
2. **T – Test**
3. **Data**
4. **Enter μ_0 =0, List: L_3, Freq:1**
5. **Choose RTT, LTT, or 2TT**
6. **Calculate**

Finding *P*-Value



***P*-value** = probability of a test statistic at least as extreme as the one obtained

p = population proportion

$\hat{p}: Sample\ Proportion$

Elementary Statistics: Correlation & Regression

The **linear correlation coefficient** $\boldsymbol{r}$, is a number that measures how well paired sample data fit a straight-line pattern when graphed . The value of r^2 is the proportion of the variation in y that is explained by the linear relationship between x and y. (**The linear correlation coefficient:** $-1 \le r \le 1$)

$$r = \frac{n\left(\sum xy\right)-\left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\left(\sum x^2\right)-\left(\sum x\right)^2\right]\left[n\left(\sum y^2\right)-\left(\sum y\right)^2\right]}}, \text{ Or}: r = \frac{\sum (Z_x Z_y)}{n-1}$$

z_x denotes the z score for an individual sample value x

z_y is the z score for the corresponding sample value y.

Regression Line:

Given a collection of paired sample data, the regression line (or line of best fit, or least-squares line) is the straight line that "best" fits the scatterplot of the data.

No significant linear correlation → The best predicted y-value is $\bar{y}$.

Significant linear correlation → The best predicted y-value is found by Plugging x-value into the regression equation.

Regression Line:

$$y = b_0 + b_1 x,$$

$$Slope: b_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

$$Y-intercept: b_0 = \bar{y} - b_1 \bar{x}$$

$$\bar{y} = \frac{\sum y}{n}, \bar{x} = \frac{\sum x}{n}$$

Test of Hypothesis:

Step 1: $H_0: \rho = 0$, $H_1: \rho \ne 0$ **claim & Tails**

Step 2: TS: $t = r\sqrt{\frac{n-2}{1-r^2}}$, **OR:** r

Step 3: CV using α ***From the T-table or Correlation (r-table)***

Step 4: Make the decision to

a. Reject or not H_0

b. The claim is true or false

c. Restate this decision: There is / is not sufficient evidence to support the claim that…

There is a linear Correlation If $|r| \ge$ **critical value**

There is No Correlation If $|r| <$ **critical value**

$$TS: t = r\sqrt{\frac{n-2}{1-r^2}}, df = n-2$$

$$Or: r$$

TI Calculator:
How to enter data:

1. **Stat**
2. **Edit**
3. **ClrList** L_1 **&** L_2
4. **Type in your data in** L_1 **&** L_2

TI Calculator:
Scatter Plot:

1. **Press on Y & clear**
2. **2nd y, Enter**
3. **On, Enter**
4. **Select X1-list:** L_1
5. **Select Y1-list:** L_2
6. **Mark: Select Character**
7. **Press Zoom & 9 to get ZoomStat**

TI Calculator:
Linear Regression - test

1. **Stat**
2. **Tests**
3. **LinRegTTest**
4. **Enter** L_1 **&** L_2
5. **Freq = 1**
6. **Choose** $\ne$
7. **Calculate**

Elementary Statistics, Goodness-of-Fit and Contingency Tables

By **"goodness-of-fit"** we mean that sample data consisting of observed frequency counts arranged in a single row or column (called a one-way frequency table) agree with some particular distribution (such as normal or uniform) being considered.

Definition: A **Multinomial Experiment** is an experiment that meets the following conditions:

1. The number of trials is fixed.
2. The trials are independent.
3. All outcomes of each trial must be classified into exactly one of several different categories.
4. The probabilities for the different categories remain constant for each trial.

The data is **randomly** selected, sample **data consist of frequency counts** for all categories & $\boldsymbol{E(x) \geq 5}$

H_0: The frequency counts agree with the claimed distribution.

H_1: The frequency counts do not agree with the claimed distribution. **RTT**

Test Statistic:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$df = k - 1$$

O = observed frequency

E = expected frequency

TI Calculator: Enter data:

1. **Stat**
2. **Edit**
3. **ClrList L_1 & L_2**
4. **O → L_1 & E → L_2**

TI Calculator:
Goodness of Fit - test

1. **Stat**
2. **Tests**
3. $\chi^2 GOF - Test$
4. **Enter L_1 & L_2**
5. **df = n − 1**
6. **Calculate**

Contingency Table: A contingency table (or two-way frequency table) is a table consisting of frequency counts of categorical data corresponding to two different variables. (One variable is used to categorize rows, and a second variable is used to categorize columns.)

Test of Independence (of variables):

The test of independence of variables is used to determine whether two variables are independent of or related to each other when a single sample is selected

Chi-Square Test of Homogeneity: (Test of test of homogeneity of proportions)

The test of homogeneity of proportions is used to determine whether the proportions for a variable are equal when several samples are selected from different populations.)

Instead of testing the null hypothesis of independence between the row and column variables, we test the null hypothesis that the different populations have the same proportion of some characteristic.

1. The data are obtained from a random sample.
2. The expected frequency for each category must be 5 or more.

H_0: The row and column variables are independent. (There is no relationship between two variables.)

H_1: The row and column variables are dependent. (There is a relationship between two variables.)

$$TS: \chi^2 = \sum \frac{(O-E)^2}{E}$$

$$df = (r - 1)(c - 1)$$

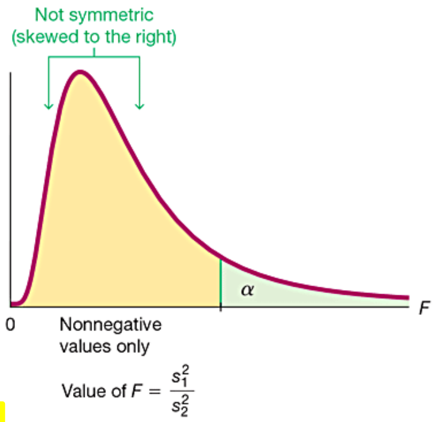
$$E = \frac{\text{(row sum)(clumn sum)}}{\text{grand total}}$$

TI Calculator:
Contingency Table

1. **Access Matrix (2nd & Press x^{-1})**
2. **Edit, Enter Dimensions & cell entries**
3. **Stat**
4. **Tests**
5. $\chi^2 - Test$
6. **Observed Matrix must be A**
7. **Calculate**

Elementary Statistics: Analysis of Variance

One-Way ANOVA



One-way analysis of variance (ANOVA) is used for tests of hypotheses that **three or more populations have means** that are all equal, as in $H_0: \mu_1 = \mu_2 = \mu_3$ by analyzing sample variances. One-way analysis of variance is used with data categorized with **one factor** (or **treatment**), so there is one characteristic used to separate the sample data into the different categories.

In the F test, two different estimates of the population variance are made.

The first estimate is called the **between-group variance**, and it involves finding the variance of the means.

The second estimate, the **within-group variance**, is made by computing the variance using all the data and is not affected by differences in the means.

If there is no difference in the means, the between-group variance will be approximately equal to the within-group variance, and the F test value will be close to 1; *do not reject null hypothesis*.

However, when the means differ significantly, the between-group variance will be much larger than the within-group variance; the F test will be significantly greater than 1; *reject null hypothesis*.

Given the means of each group:

Calculate their mean and variance:

$$\bar{x} = \frac{\sum x}{k}$$

$$s_{\bar{x}}^2 = \frac{\sum (x - \bar{x})^2}{k - 1} =$$

$$S_B = \sigma_{\bar{x}} = \sigma / \sqrt{n} \rightarrow Estimate : \sigma^2 = n s_{\bar{x}}^2$$

$$s_W^2 \ \ OR \ \ s_p^2 = \frac{S_1^2 + S_2^2 + S_3^2}{k}$$

$$TS : F = \frac{s_B^2}{s_W^2}$$

TI Calculator:

One Way ANOVA

1. **ClrList $L_1, L_2, \ldots$**
2. **Enter data $L_1, L_2, \ldots$**
3. **Stat**
4. **Tests**
5. ***ANOVA*** $(L_1, L_2, \ldots)$
6. **Enter**

Given:

*Number of Groups (**Factors**): k*

Number of data in each group: *n*

Total sample size: N = k(n)