Chapter 1 Study Guide: Exploring Data

Ms. Kempton

Individuals: The objects described by a set of data. (*ex: people, animals, things…*)

Variable: Any characteristic of an individual.

Categorical Variable: Places an individual into groups or categories. (*ex: colors*)

Quantitative Variable: Variables that are numbers. (*ex: height*, *weight*)

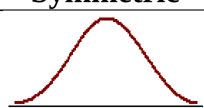
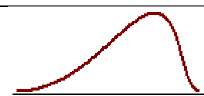
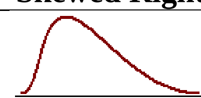
<u>Describe Distribution</u>:

- **S**hape: symmetric or skewed
- **O**utliers: value that is outside pattern
- **C**enter: mean/median
- **S**pread(Variability): range

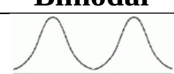
} + CONTEXT

<u>Compare Distributions</u>:
Use comparative words
(similar, greater than, less than)

<u>SHAPE of a distribution</u>: Use –ly words (slightly, moderately, strongly)

Symmetric	Skewed Left	Skewed Right
		
Mean = Median	Mean < Median	Mean > Median

Other words to describe shape:

Bimodal	Uniform
	

<u>CENTER of a distribution</u>:

~**Mean:** → Use with symmetric data

~**Median:** middle point of a distribution$\left(location = \frac{n+1}{2}\right)$ → Use with skewed data

<u>SPREAD/VARIABILITY of a distribution</u>:

~**Range** = max – min

~ **Standard Deviation** = $\sqrt{variance}$ → The average distance from the mean

→ The <u>(*context*)</u> typically varies by <u>(*SD*)</u> from the mean by <u>($\bar{x}$)</u>.

~**IQR** = IQR = $Q_3 - Q_1$

Resistant: A measure that is unaffected by extreme values (ex-median)

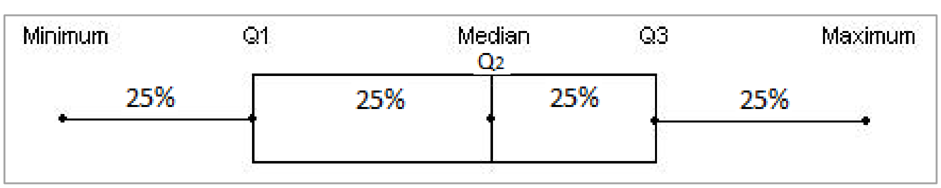
~Skewed data or outliers → Use median and IQR

~Symmetric data → Use mean and standard deviation

<u>5 Number Summary</u>: MIN Q_1 MED Q_3 MAX → Use to make boxplot

Outliers: way too small < $Q_1 - 1.5IQR$ and way too big > $Q_3 + 1.5IQR$

Boxplots:

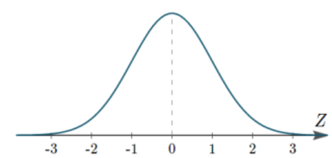


Chapter 2 Study Guide:
Modeling Distributions of Data

Percentile: The value that has p% of data less than or equal to it

Z-score: The number of standard deviations above or below the mean.

$$z = \frac{x - mean}{standard\ deviation}$$

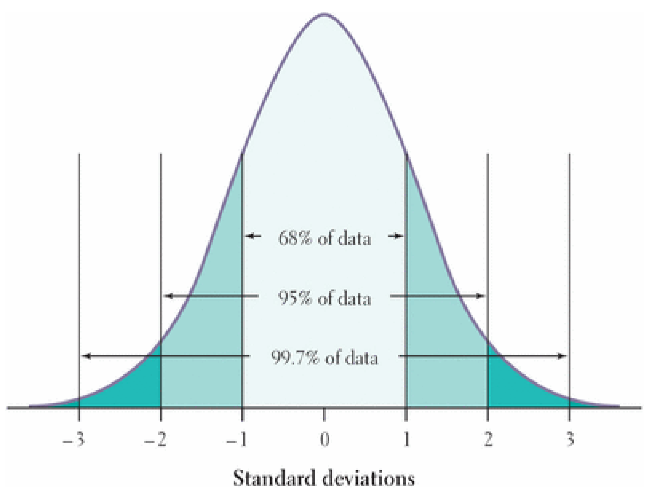


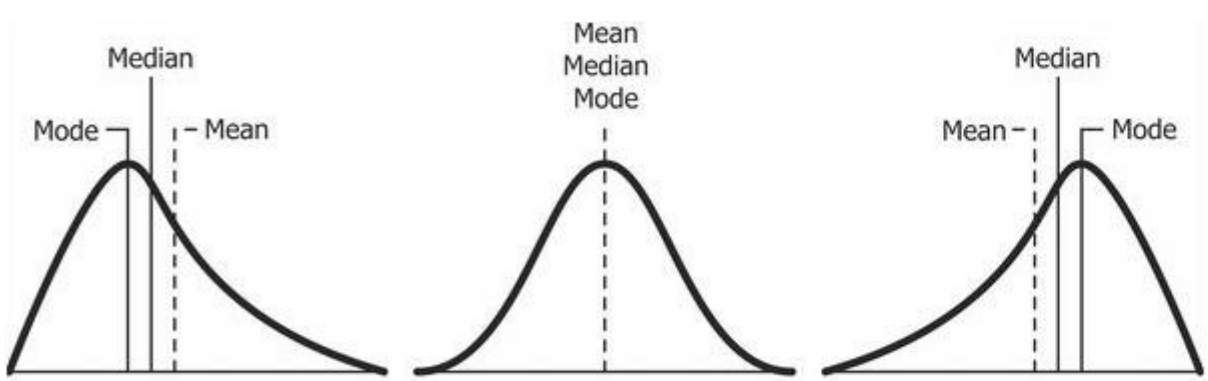
Transforming Data:

	+/−	÷/×
S (shape)	S	S
C (center)	C	C
S (spread)	S	C

Density Curve: Area under the curve is 1 (100%).

Normal Distribution: Symmetric and bell-shaped



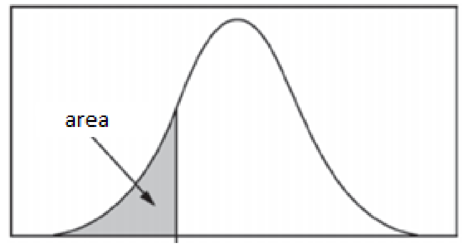


Skewed to the Right **Symmetric** **Skewed to the Left**

Area under the curve:

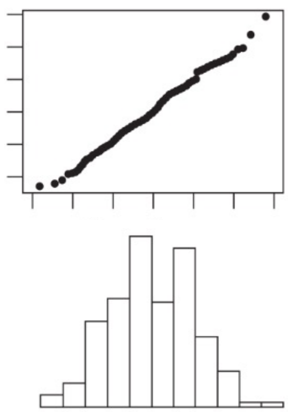
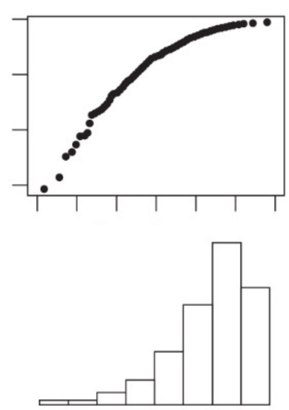
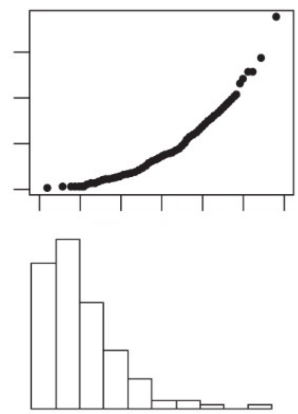
TO FIND AREA ⟶ 2^{nd} → VARS → normalcdf(
Use upper and lower as -10000 or 10000

IF GIVEN AREA ⟶ 2^{nd} → VARS → invNorm(
~Only finds area to the left

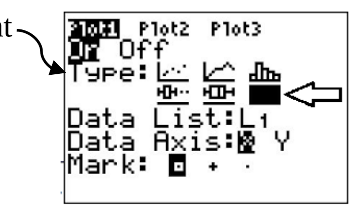


Normal Probability Plot:

If the points are close to a straight line, then the data are approximately **Normally distributed**.

Approximately Normal	Skewed to the Left	Skewed to the right
		

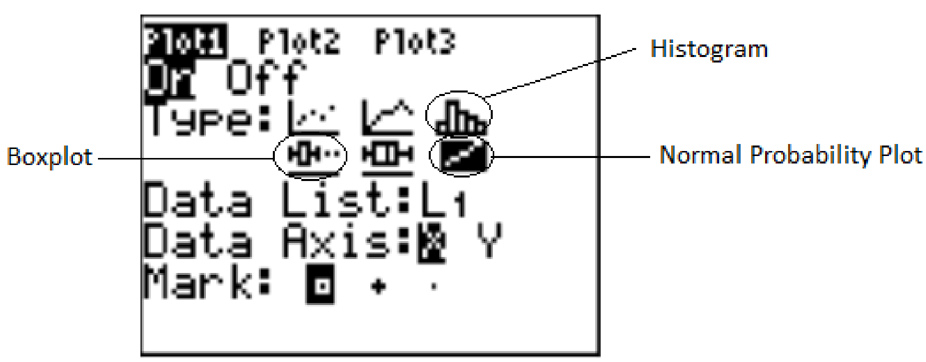
1) **Turn Plot1 On:** 2^{nd} → STAT PLOT → ENTER → On
2) **Set Type:** Use arrow keys to highlight the Type to bottom right
3) **Graph:** ZOOM → 9



Does your graph look like an approximate straight line? _______

*** If the points of the NORMAL probability plot are close to a straight line, then the data are approximately **Normally distributed**.

Graphing a Histogram, Boxplot, and Normal Probability Plot



Chapter 3 Study Guide:
Exploring Two-Variable Data

Ms. Kempton

Explanatory Variable: Used to predict. (*x*-values)

Response Variable: What you want to predict. Outcome, responds to explanatory. (*y*-values)

Describe Relationship: (DUFS + context)

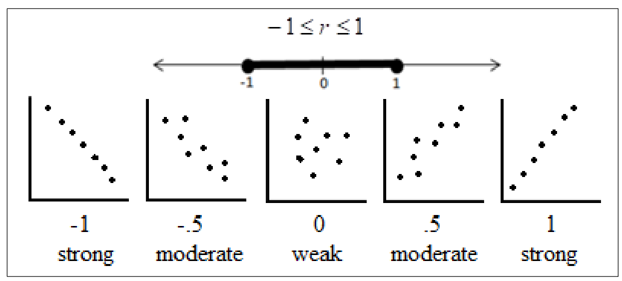
Direction: +/ –
Unusual points: possible outliers
Form: linear
Strength: strong/moderate/weak

} + CONTEXT

<u>Template for Describing Relationship & Interpreting Correlation(*r*)</u>:

There is a ____________, ____________ <u>linear</u> relationship between ____________ and ____________.
(strength) (direction)
ly words

Correlation (*r*): Relationship/Association
~Direction: +/ –
~Strength: strong/moderate/weak
(use –ly words for in between)



<u>Facts/Cautions about Correlation</u>:

- *r* doesn't have units → it is just a number used to measure direction and strength!
- Switching axes doesn't change *r* → any variable can be *x* and *y*
- *r* nonresistant → strongly affected by outliers & high-leverage points
  ~ high-leverage points in pattern strengthen
  ~outliers out of pattern weaken
- **Correlation does NOT imply causation**!!!

Outlier: A point that does not follow the pattern and has a large residual. (far up/down)

High-Leverage: A point that has a much larger or smaller *x* value than the other points. (far left/right)

Influential Point: A point that, if removed, changes the relationship (slope/*y*-int/*r*/r^2, or s) substantially.

*<u>Outliers</u> and <u>high-leverage</u> points are often <u>influential</u>!

<u>Least-Squares Regression Line</u>: used to describe and/or predict the relationship between two variables.

$$\hat{y} = a + bx$$

predicted y-int slope

~Interpreting *y*-intercept: *When* ___________, *the predicted* ___________ *is* ___________.
(*x* context = 0) (*y* context) (*y*-intercept)

~Interpreting Slope: *For every additional* ________ *, the predicted* ________ *increases/decreases by* _______.
(units in x context) (*y* context) (slope)

~Extrapolation: using the LSRL to make a prediction far outside the observed x values.

<u>Residual</u>: Actual *y* – Predicted $\hat{y}$ $(R = A - P)$

~Interpreting Residual: *The actual* ___________ *is higher/lower than the predicted by* ___________.
(y-context) (residual)

<u>Standard Deviation (*s*) of the Residuals</u>: used to assess how well the line fits all the data. This value gives the approximate size of a "typical" prediction error (residual).

- Interpretation: *The actual <u>(y-context)</u> is typically about <u>(standard deviation)</u> away from the predicted <u>(y-context)</u> by the LSRL.*

<u>Coefficient of Determination</u> (r^2): the *fraction/percent* (or piece) of the variation in the values of y that can be explained by the least-squares regression line.

- Interpretation: *About* _____*% of the variability in <u>(y-context)</u> can be explained by the LSRL.*

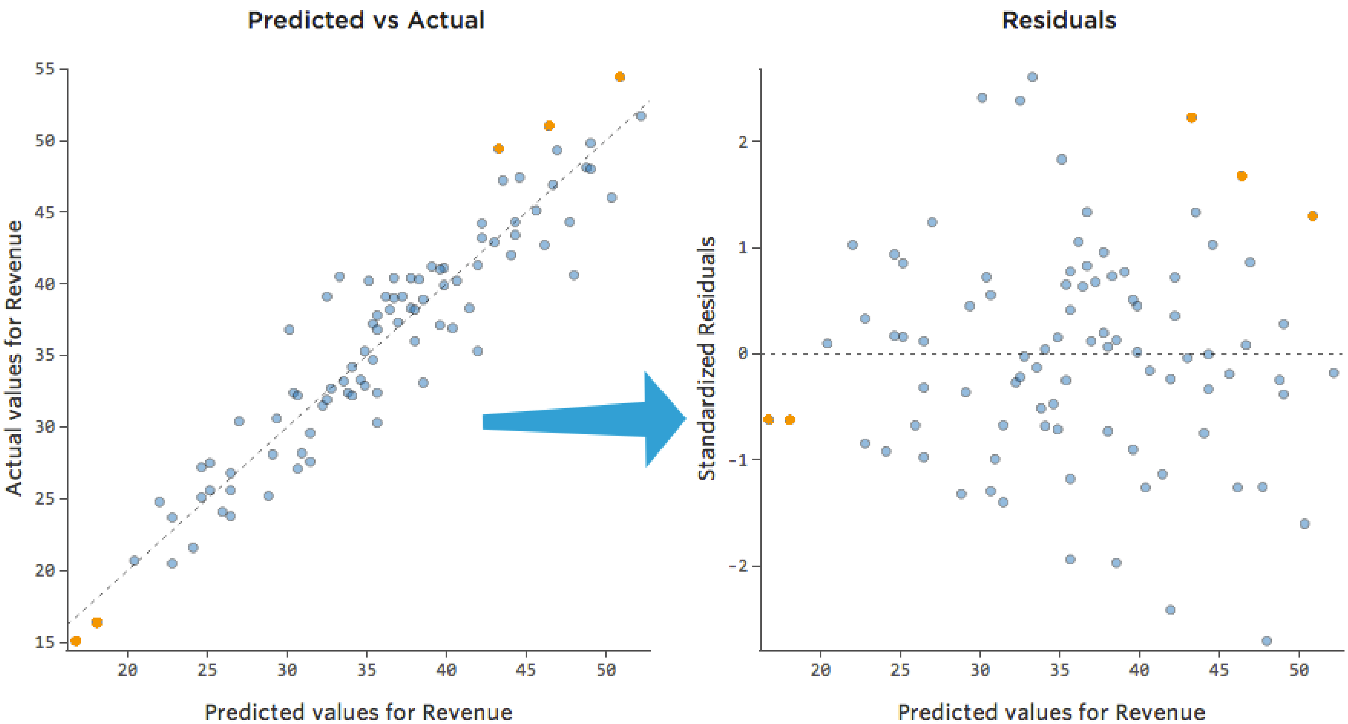
y-intercept (a)
~When ___(x=0), the predicted ___(y) is ___.

slope (b)
~For every one ___(unit in x), the predicted ___(y) increases/decreases by ___.

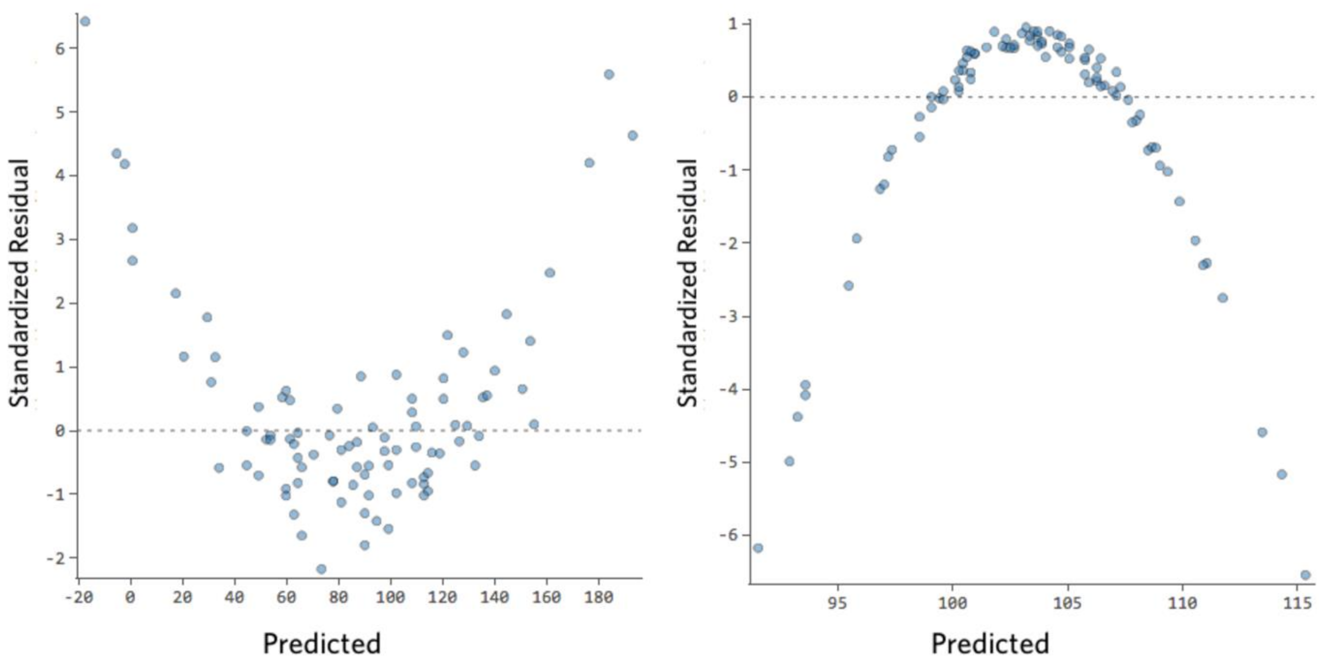
Predictor	Coef	SE Coef	T	P
Constant	1.201	~~0.0874~~	~~13.72~~	~~0~~
GPA **(X)**	7.507	~~1.29~~	~~5.82~~	~~0.0006511~~
S = 3.252686	R-Sq = 82.8%		~~R-Sq(adj) = 76.5%~~	

SD of residuals (s)
~The actual ___(y) is typically about ___(SD) away from the predicted ___(y) by the LSRL.

Coefficient of Determination (r^2)
~___% of the variation in ___(y) can be explained by the LSRL.


Predicted vs Actual
Actual values for Revenue
Predicted values for Revenue
Residuals
Standardized Residuals
Predicted values for Revenue

Nonlinear


Standardized Residual
Predicted
Standardized Residual
Predicted

Chapter 4 Study Guide: Designing Studies

Ms. Kempton

Population: All items or subjects.

Sample: Subset of population

Census: Selection of all items/subjects in a population

Sampling Methods (DATA COLLECTION):

- **Simple Random Sample (SRS):** Every group of size *n* has an equal chance of being chosen.

Describing a SRS Procedure:	
Hat Trick	Technology
1) **Label** individuals on equal-sized pieces of paper. (or assign each individual a number) (or assign each individual a number) 2) **Randomize**: Put them in a hat and mix well. 3) **Select:** Randomly select _____ (individuals) (or those corresponding to the numbers) for the sample. Mix thoroughly after each selection. *Individuals: Who/what you are sampling	1) **Label** each individual with a number. 2) **Randomize:** Using a random number generator, generate _____ random numbers from _____ to _____, ignoring repeats. 3) **Select:** The (individuals) corresponding to those numbers would be selected for the sample. *In Calculator: Math → PRB → randInt(1,n)

- **Stratified Random Sample:** Split the population into groups of similarity (strata), then take an SRS of each group. Those chosen from each SRS will be selected for the sample.
  - **Homogeneous grouping**
- **Cluster** Split population into groups based on location (clusters), then randomly select clusters (1 SRS). All individuals within the cluster are selected for the sample.
  - **Heterogeneous grouping**
- **Systematic Random Sample:** Randomly select a starting point and every *n*th individual thereafter will be selected for the sample.

Non-Random Sampling Methods:

- **Convenience Sample:** Choosing individuals easy to reach. (Bias)
- **Voluntary Response Sample:** People who choose themselves to be in the sample. (Bias)
  - Using people with strong opinions

<u>Bias</u>: Favoring a group (not fair).

- **Undercoverage Bias:** When members of the population are less likely to be chosen for a sample
- **Nonresponse Bias:** Individuals selected to be in a sample but can't be contracted or refuse.
- **Response Bias:** Inaccurate responses (lying or confusing questions).

<u>Observational Study</u>: No treatments imposed, only observed. (NO CAUSE/EFFECT)

- **Retrospective:** examine existing data
- **Prospective:** track individuals into the future

<u>Experiment</u>: Treatments imposed

- **Experimental Units:** Individuals assigned treatments. (Humans often called **subjects**)
- **Explanatory Variable:** Variable whose levels are manipulated intentionally.
- **Response Variable:** Outcome from the treatments administered.
- **Confounding Variables:** Variables that influence the response variable.

<u>Well-Designed Experiment</u>:

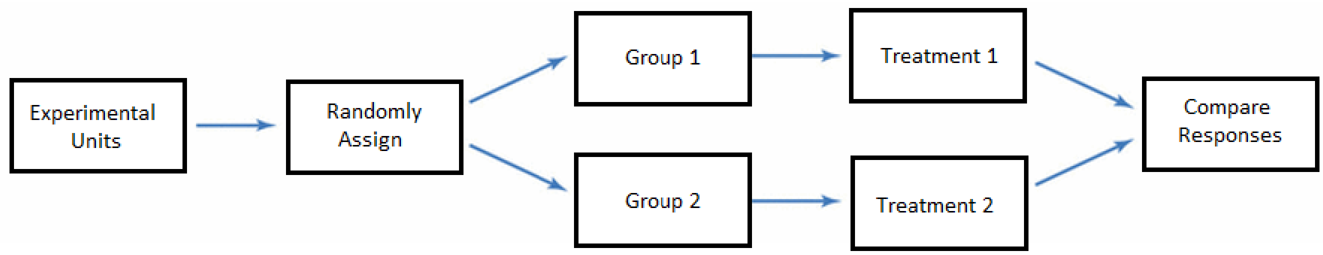
a) **Comparison:** Compares at least two treatment groups.

b) **Random Assignment:** Experimental units randomly assigned to treatments (balances confounding variables).

c) **Control:** Control potential confounding variables.

d) **Replication:** Enough experimental units in each treatment group (more than one).

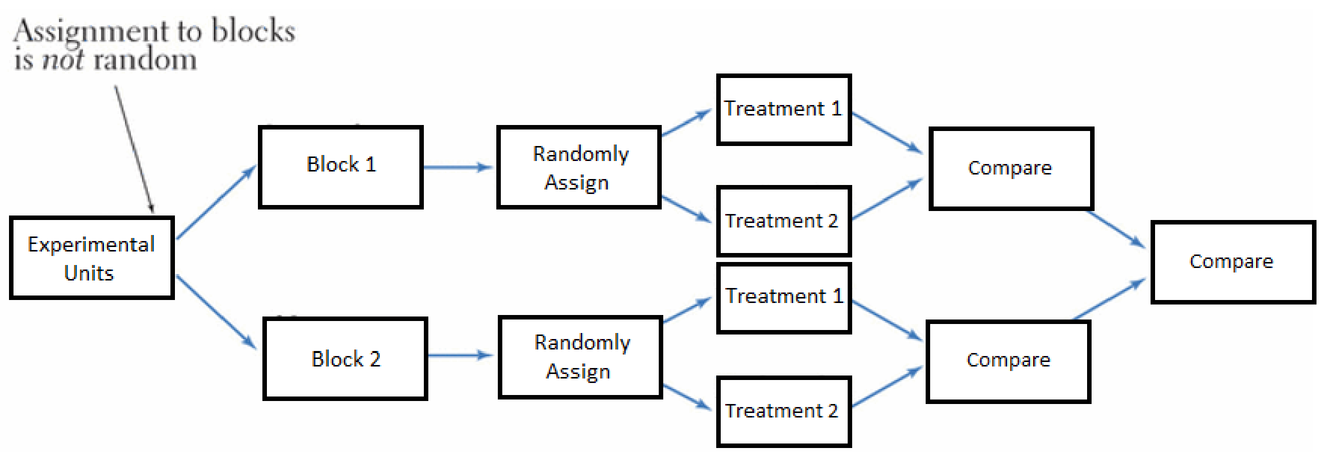
Completely Randomized Design:



- **Random Selection:** Generalizes population.
- **Random Assignment:** Allows causation.
- **Single-blind Experiment:** Subjects do not know which treatment they are receiving.
- **Double-blind Experiment:** Neither the subjects nor the experimenters who interact with the subjects know which treatment a subject is receiving.
- **Control Group:** A group receiving an inactive treatment (**Placebo**).
- **Placebo Effect:** When experiment units have a response to a placebo (fake treatment works).

Randomized Block Design: At the beginning of the experiment, units are divided.

- **Block:** Group of experimental units that are similar in some way.



Matched Pairs Design: A pair or experimental units that are matched based off similarity then randomly assigned to each treatment.

- Sometimes a "pair" consists of a single unit that receives both treatments. The order of the treatments is randomly chosen.

Statistically Significant: When results from a study are too unusual to have occurred purely by chance.

- Proportion of dots are $\leq 5\%$ → STATISTICALLY SIGNIFICANT

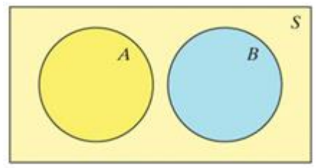
Chapter 5 Study Guide: Probability

Ms. Kempton

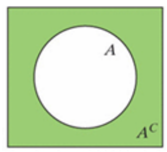
Law of Large Numbers: States that when an experiment is performed a large number of times, the relative frequency (proportion) of an event tends to become closer to the actually probability.

Probability Model: List showing all possible outcomes and their probabilities.
- Each probability must between 0 and 1
- All probabilities (sample space) must add to 1

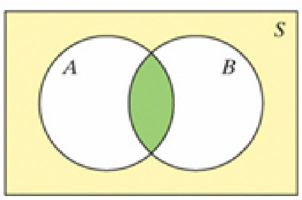
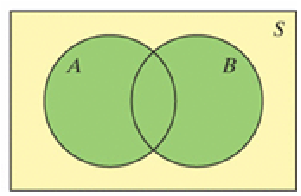
Mutually Exclusive: Events that cannot occur together
- If events are mutually exclusive → $P(A \text{ or } B) = P(A) + P(B)$



Complement Rule: $P(A^C) = 1 - P(A)$



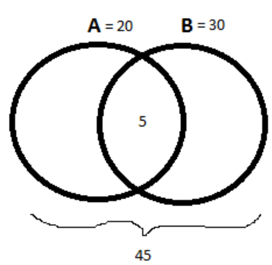
<u>Other Venn Diagrams and Probability Rules:</u>

A and B ($A \cap B$) "both"	**A or B ($A \cup B$)** "either"
	

General Addition Rule: (ON FORMULA SHEET)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Example: $P(A \text{ or } B) = 20 + 30 - 5 = 45$



Given:

$P(A|B)$ → (<u>out of</u> B's, how many A's?)

Independent: One event has no effect on the probability of the other event

$A|B = A$

If events are independent → $P(A \text{ and } B) = P(A) \cdot P(B)$

Chapter 6 Study Guide: Random Variables

Ms. Kempton

Random Variable: Variable whose value is a numerical outcome of some chance process.

Discrete Random Variable: A random variable that takes a fixed set of possible values with gaps between.

Probability Distribution:

Value:	x_1	x_2	x_3	...
Probability:	p_1	p_2	p_3	...

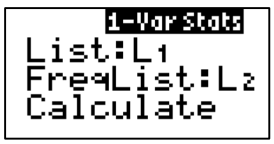
To find Expected Value (mean): Think "multiply and add"

Expected value (mean) = $x_1 p_1 + x\ p_2 + x_3 p_3 + ... = \sum x_i p_i$ → <u>ON FORMULA SHEET</u>:

$$\mu_X = E(X) = \Sigma x_i P(x_i)$$

*You can also find mean and SD in calculator:

1) stat → Edit
2) Input values into L_1 and L_2
3) stat → CALC → 1-VarStats ⟶



Continuous Random Variable: A random variable that takes all values in an interval of numbers.

- To find continuous random variable probabilities → Use normalcdf (2^{nd} → VARS → normalcdf)

<u>Transforming and Combining Random Variables</u>:

TRANFORMATION RULES:

<u>Adding/Subtracting a Constant</u>:

- <u>Same</u>: **shape** or **spread** (range, IQR, standard deviation)
- <u>Change</u>: **center** (mean, median, quartiles, percentiles)

<u>Multiplying/Dividing a Constant</u>:

- <u>Same</u>: **shape**
- <u>Changes</u>: **center** (mean, median, quartiles, percentiles) or **spread** (range, IQR, standard deviation)

	+/−	×/÷
S	S	S
C	C	C
S	S	C

Mean of the Sum/Difference of Random Variables

$$\mu = \mu \pm \mu$$

*The mean of the sum of *several* random variables is the <u>sum or difference</u> of their means.

Variance of the Sum/Difference of Independent Random Variables

$$\sigma^2 = \sigma^2 + \sigma^2$$

*The variance of the sum of *several independent* random variables is the <u>sum</u> of their variances

Binomial Probability:

Binomial Setting: (BINS)

- **B**inary: “success” or “failure”
- **I**ndependent: One trial’s outcome does not affect any other trial
- **N**umber: Set number of trials *n* = ____
- **S**uccess: Same probability *p* = ____

To find an EXACT probability → use **binompdf**

- ➢ 2nd → VARS → binompdf(

```
binompdf
trials:
p:
x value:
Paste
```

To find a NO MORE THAN probability → use **binomcdf**

- ➢ 2nd → VARS → binompcf(

To find an AT LEAST probability → use **binomcdf**

- ➢ 2nd → VARS → 1– binompcf(

```
binomcdf
trials:
p:
x value:
Paste
```

Mean and Standard Deviation of a Binomial Distribution: ON FORMULA SHEET:

Mean → $\mu_X = np$ Standard Deviation: $\sigma_X = \sqrt{np(1-p)}$

Geometric Probability:

Geometric Setting: (BINS without the N)

- **B**inary: “success” or “failure”
- **I**ndependent: One trial’s outcome does not affect any other trial
- **S**uccess: Same probability *p* = ____

To find an EXACT probability → use **geometpdf**

- ➢ 2nd → VARS → geometpdf(

```
geometpdf
p:
x value:
Paste
```

To find a NO MORE THAN probability → use **geometcdf**

- ➢ 2nd → VARS → geometcdf(

To find an AT LEAST probability → use **binomcdf**

- ➢ 2nd → VARS → 1– geometcdf(

```
geometcdf
p:
x value:
Paste
```

Mean and Standard Deviation of a Geometric Distribution: ON FORMULA SHEET:

Mean → $\mu_X = \frac{1}{p}$ Standard Deviation: $\sigma_X = \frac{\sqrt{(1-p)}}{p}$

Chapter 7 Study Guide: Sampling Distributions

Ms. Kempton

Sampling Distribution: The distribution of values from all possible samples of size n.

Parameter: A number that describes a **population**.

- $p =$ population proportion and $\mu =$ population mean

Statistic: A number that describes a **sample**.

- $\hat{p} =$ sample proportion and $\bar{x} =$ sample mean

Unbiased Estimator: The mean of the sampling distribution is equal to the population parameter.

- $\hat{p} = p$ and $\bar{x} = \mu$
- A larger sample size will decrease variability (**L**arger sample = **L**ess variability)

Proportions:	**Means**:
One Population → $\mu_{\hat{p}} = p$ $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$	One Population → $\mu_{\bar{x}} = \mu$ $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$
Two Populations → $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$ $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$	Two Populations → $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$ $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\dfrac{\sigma_1{}^2}{n_1} + \dfrac{\sigma_2{}^2}{n_2}}$
Conditions: **Independent:** $n \le 10\%N$ **Approximately Normal →** $np \ge 10$ $n(1-p) \ge 10$	**Conditions**: **Independent:** $n \le 10\%N$ **Normal →** Population Normal **Approximately Normal →** $n \ge 30$

Chapter 8 Study Guide: Confidence Intervals

Ms. Kempton

Statistic $\pm$ (critical value)(standard error of statistic)
(*point estimate*)

One-Sample z Interval for a Population Proportion

In Calculator:	By Hand:
STAT → TESTS → 1-PropZint (A) → **1-PropZInt** / **x:** / **n:** / **C-Level:**	$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ Find z: 2nd → VARS → **invNorm(** / **area:** / μ: / σ:

One-Sample *t* Interval for a Population Mean: (When Population SD is unknown)

In Calculator:	By Hand:
STAT → TESTS → TInterval (8) → **TInterval** / **Input: Stats** / $\bar{x}$: / **Sx:** / **n:** / **C-Level:**	$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$ Find *t*: 2nd → VARS → **invT(** / **area:** / **df:** (**df** = n-1)

Width of confidence interval:
-Confidence interval ↓ as sample size *n* ↑
-Confidence interval ↑ as confidence level ↑

statistic ± (critical value)(standard error of statistic)
(critical value)(standard error of statistic) = ME

To find sample size: $z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq ME$

- *If $\hat{p}$ is not given → use 0.5*

Interpreting Confidence Level:

*In repeated random samples of size _____(n) from this population, about _____% of confidence intervals created will capture the true population (proportion or mean) of…

*KEMP DEF: *In the long run, _____ % of intervals generated capture the population proportion/mean*

Confidence Intervals: A Four-Step Process

	Proportions	**Means**
S T A T E	p = true **proportion** of… Confidence Level = ______	μ = true **mean** of… Confidence Level = ______
P L A N	<u>Name</u>: One Sample z interval for p <u>Conditions</u>: 1) Random sample 2) Independent if $n \le 10\% N$ 3) App Normal if $n\hat{p} \ge 10$ and $n(1-\hat{p}) \ge 10$	<u>Name</u>: One Sample *t* interval for μ <u>Conditions</u>: 1) Random sample 2) Independent if $n \le 10\% N$ 3) Approximately Normal if: ➢ $n \ge 30$ ➢ Sample shows no strong skewness and outliers
D O	$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ OR 1-PropZInt (A)	$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$ OR TInterval (8)
C O N C L U D E	We are _____% confident that the interval from _____ to _____ captures the true **proportion** of…	We are _____% confident that the interval from _____ to _____ captures the true **mean** of…

Chapter 9 Study Guide: Significance Tests

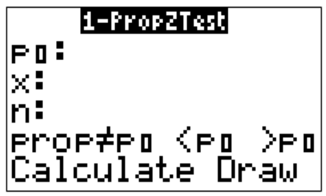
Ms. Kempton

$$\textbf{Standardized test statistic} = \frac{\textbf{statistic} - \textbf{parameter}}{\textbf{standard error of the statistic}} = \frac{\boldsymbol{stat - null}}{\boldsymbol{SD}}$$

One-Sample z Test for a Population Proportion

In Calculator:

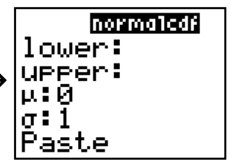
STAT → TESTS → 1-PropZTest (5) →



By Hand:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

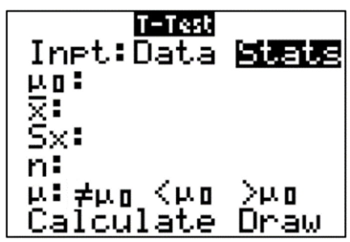
2nd → VARS →



One-Sample *t* Test for a Population Mean

In Calculator:

STAT → TESTS → T-Test (2) →



By Hand:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

2nd → VARS →

(**df** = n-1)

Errors (HOT HAT):

- **Type I Error (H_0t):** Reject H_0 when H_0 is true
- **Type II Error (H_at):** Fail to reject H_0 when H_a is true

	H_0 true	H_a true
Reject H_0	Type I error α	Correct conclusion → POWER
Fail to reject H_0	Correct conclusion	Type II error β

Power: Probability that the test will correctly reject H_0

$$P = 1 - \beta$$

(β → prob of type II error)

P-value: Assuming the H_0 is true, the probability of one sample occurring

α: Probability of making a type I error

β: Probability of making a type II error

If reject H_0
- Correct decision → POWER
- Wrong decision → TYPE I ERROR

Significance Tests: A Four-Step Process

	Proportions	Means
STATE	p = true **proportion** of… H_0: p = H_a: p ≠, <, or >	μ = true **mean** of… H_0: μ = H_a: μ ≠, <, or >
PLAN	<u>Name</u>: One Sample z Test for p <u>Conditions</u>: 1) Random sample 2) Independent if $n \le 10\%N$ 3) App Normal if $np_0 \ge 10$ and $n(1-p_0) \ge 10$	<u>Name</u>: One Sample *t* Test for μ <u>Conditions</u>: 1) Random sample 2) Independent if $n \le 10\%N$ 3) Approximately Normal if: ➢ $n \ge 30$ ➢ Sample shows no strong skewness and outliers
DO	$z = \dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ OR 1-PropZTest (5)	$t = \dfrac{\bar{x}-\mu_0}{\dfrac{s}{\sqrt{n}}}$ OR T-Test (2)
CONCLUDE	p-value ≤ α , reject H_0. We have convincing evidence that the true **proportion** of…(H_a) p-value > α, fail to reject H_0. We do not have convincing evidence that the true **proportion** of…(H_a)	p-value ≤ α , reject H_0. We have convincing evidence that the true **mean** of…(H_a) p-value > α, fail to reject H_0. We do not have convincing evidence that the true **mean** of…(H_a)

Chapter 10 Study Guide:
Comparing Two Populations or Groups

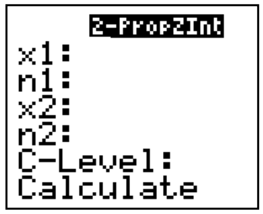
Ms. Kempton

Proportions

Two Sample z Interval for $p_1 - p_2$

In Calculator:

STAT → TESTS → 2-PropZint (B) →

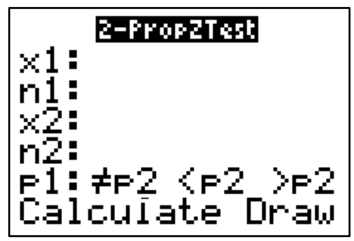


By Hand

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Two Sample z Test for $p_1 - p_2$

In Calculator: STAT → TESTS → 2-PropZTest (6) →

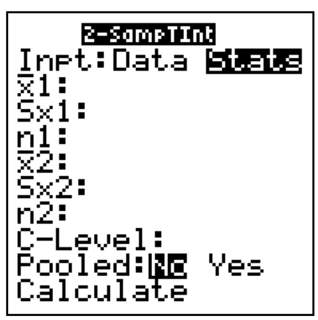


Means

Two Sample t Interval for $\mu_1 - \mu_2$

In Calculator:

STAT → TESTS → 2-SampTInt (0) →

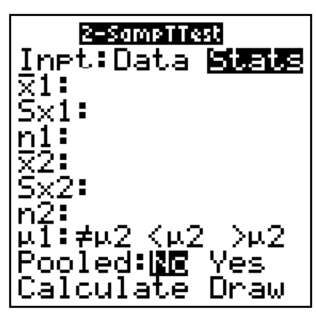


By Hand

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Two Sample t Test for $\mu_1 - \mu_2$

In Calculator: STAT → TESTS → 2-SampTTest (4) →



Confidence Intervals: A Four-Step Process

	Proportions	Means
S T A T E	$p_1 - p_2$ = true difference in **proportions** of… Confidence Level = ______	$\mu_1 - \mu_2$ = true difference in **means** of… Confidence Level = ______
P L A N	Name: Two Sample z interval for $p_1 - p_2$ Conditions: 1) Random samples 2) Independent if: $n_1 \le 10\%N$ $n_2 \le 10\%N$ 3) Approximately Normal if: $n_1\hat{p}_1 \ge 10$ and $n_1(1-\hat{p}_1) \ge 10$ $n_2\hat{p}_2 \ge 10$ and $n_2(1-\hat{p}_2) \ge 10$	Name: Two Sample *t* interval for $\mu_1 - \mu_2$ Conditions: 1) Random samples 2) Independent if: $n_1 \le 10\%N$ $n_2 \le 10\%N$ 3) Approximately Normal if: ➢ $n_1 \ge 30$ and $n_2 \ge 30$ ➢ Sample shows no strong skewness and outliers
D O	$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ OR 2-PropZint (B)	$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ OR 2-SampTInt (0)
C O N C L U D E	We are _____% confident that the interval from ______ to ______ captures the true difference in **proportions** of…	We are _____% confident that the interval from ______ to ______ captures the true difference in **means** of…

Significance Tests: A Four-Step Process

<table>
<tr><th></th><th>Proportions</th><th>Means</th></tr>
<tr><td>S
T
A
T
E</td><td>H$_0$: $p_1 = p_2$
H$_a$: ≠, <, or >

p_1 = true proportion of…
p_2 = true proportion of…</td><td>H$_0$: $\mu_1 = \mu_2$
H$_a$: μ ≠, <, or >

μ_1 = true mean of…
μ_2 = true mean of…</td></tr>
<tr><td>P
L
A
N</td><td><u>Name</u>: Two Sample z Test for $p_1 - p_2$

<u>Conditions</u>:
1) Random samples
2) Independent if:
$n_1 \leq 10\%N$
$n_2 \leq 10\%N$
3) Approximately Normal if:
$n_1\hat{p}_1 \geq 10$ and $n_1(1-\hat{p}_1) \geq 10$
$n_2\hat{p}_2 \geq 10$ and $n_2(1-\hat{p}_2) \geq 10$</td><td><u>Name</u>: Two Sample t Test for $\mu_1 - \mu_2$

<u>Conditions</u>:
1) Random samples
2) Independent if:
$n_1 \leq 10\%N$
$n_2 \leq 10\%N$
3) Approximately Normal if:
➢ $n_1 \geq 30$ and $n_2 \geq 30$
➢ Sample shows no strong skewness and outliers</td></tr>
<tr><td>D
O</td><td>2-propZTest (6)</td><td>2-SampTTest (4)</td></tr>
<tr><td>C
O
N
C
L
U
D
E</td><td>p-value ≤ α , reject H$_0$. We have convincing evidence that…(H$_a$)

p-value > α, fail to reject H$_0$. We do not have convincing evidence…(H$_a$)</td><td>p-value ≤ α , reject H$_0$. We have convincing evidence that…(H$_a$)

p-value > α, fail to reject H$_0$. We do not have convincing evidence that…(H$_a$)</td></tr>
</table>