**Chapter 11 Statistics – (Study) Formula Sheet**

**11.1 – Statistical Studies**

Graphical user interface

Description automatically generated**Chart, scatter chart, box and whisker chart

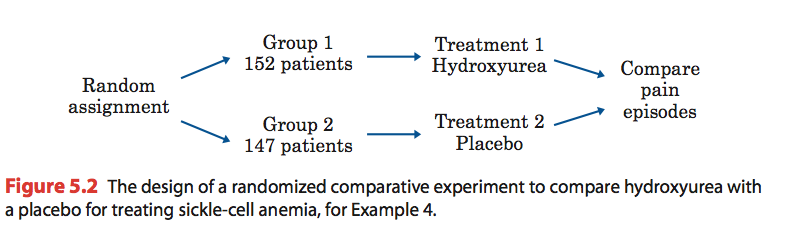
Description automatically generated**

Sampling techniques

Observational Study vs Experiment

* **Observational study** – Observes existing data.
  + Can reveal association or correlation between variables, but not causation.
* **Experiment** – Generates data to help identify cause-and-effect relationships.
  + Imposes treatments and controls randomly to groups.

Principles of Experimental Design

1. ****Randomize the control and treatment groups.
2. Control for outside effects on the variable.
3. Replicate the experiment a significant number of times to see meaningful patterns.

**11.2 – Displaying Data**

Frequency Tables

Examples

a) What percent of observations have between 1 and 4 pets inclusive?

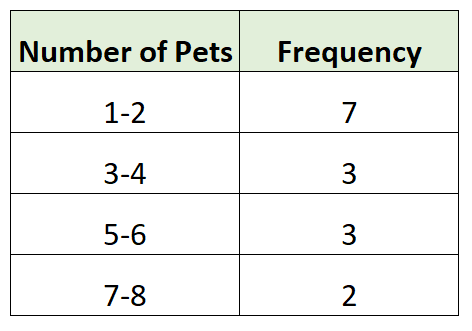
b) What percent of students prefer Football or Hockey?

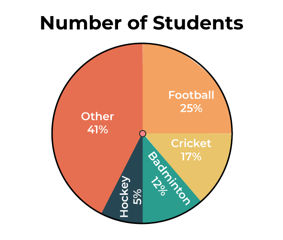
* Summarize datasets by counting the number of observations for

each category, distinct value or interval.

Table

Description automatically generated





Graphical Displays of Data

* Pie charts (categorical data)
  + Compare parts to a whole

(slices are proportion of a category).

c) Bar graph – Which season has the highest frequency?

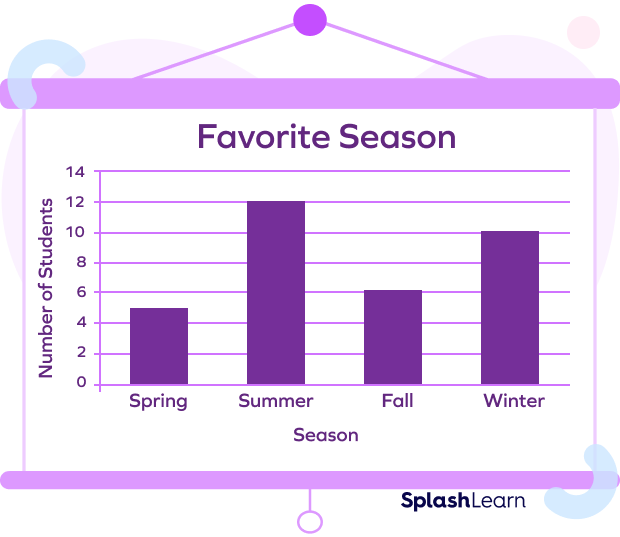
d) Histogram – How many items cost between $11 and $40 inclusive?

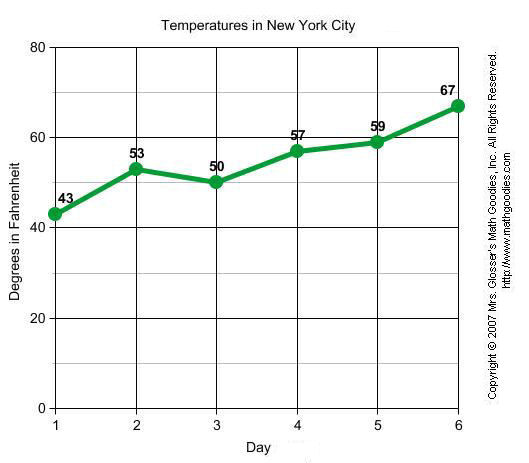
e) How many days was the temperature between 55 and 60 °F?

* Bar graphs (categorical data) and Histograms (numeric data)
  + Height of bar represents amount of data

in each category (counts or relative frequencies).

Chart, histogram

Description automatically generated

* Line graph
  + Shows changes in a numerical

variable over time.

**11.3 – Describing and Analyzing Data**

Example

Dataset: 1, 2, 7, 3, 6, 9, 1, 0, 4, 7

a) Find the mean. **Calc: 1-Var Stats**

(Data in L­1)

b) Find the median.

c) Find the mode.

d) Find the range.

e) Find the sample standard deviation.

Measures of Center

* **Mean** (average) =
  + NOT resistant 🡪 Affected by outliers
* **Median** (middle)
  + The middle value in an ordered list.
  + Resistant 🡪 NOT affected by outliers.
* **Mode** (most common)
  + The most frequently occurring value(s).
  + Resistant 🡪 NOT affected by outliers.
  + Only measure of center that can be used with categorical data.

Measures of Spread

* **Range** = Max – Min
* **Standard deviation**
  + Measures average distance from the mean.
  + (Don’t calculate by hand).

A graph with arrows and a few words

Description automatically generated with low confidenceMeasures of Relative Position

* A **percentile** tells you the percent of observations/individuals you are higher than.
  + **Quartiles** are specific percentiles.
    - A diagram of a diagram

      Description automatically generatedQ1 is the 25th Percentile.
  + Q3 is the 75th Percentile.
    - Q2 is the 50th Percentile = Median.
* **Inner Quartile Range (IQR)**
  + IQR = Q3 – Q1
* **5-number summary**
  + Min, Q1, Med, Q3, Max 🡪 Points of a boxplot
* Example: Calculate the 5-number summary and sketch a boxplot for the following dataset.
  + 12, 3, 4, 7, 21, 3, 9, 8, 10, 11, 25, 11, 13, 4, 5

**11.4 – The Normal Distribution**

Chart, histogram

Description automatically generatedEmpirical Rule (68 – 95 – 99.7 Rule)

68% of the data lies within 1 st dev of the mean.

95% of the data lies within 2 st devs of the mean.

99.7% of the data lies within 3 st devs of the mean.

* Finding probabilities using the Empirical Rule.

Example

Oak tree heights are normally distributed with

mean 60 m and st dev 5 m.

a) Find the percent of trees between 50 m and 70 m tall.

b) Find the percent of trees greater than 65 m

* + Step 1 🡪 **Draw** and **label** curve.
  + Step 2 🡪 **Shade** curve.
  + Step 3 🡪 **Use empirical rule**.

A picture containing text, lamp

Description automatically generated

A picture containing text, lamp

Description automatically generated

Diagram

Description automatically generatedFinding probabilities based on the normal distribution

* Step 1 🡪 **Standardize** using the **z-score**.
  + Formula:
  + Ex) *X* has a normal distribution with mean 10 and st dev 2.

Find the z-score for .

Diagram

Description automatically generated with medium confidence

* Step 2 🡪 **Draw**, **label** and **shade** curve.
  + This is how you show your work!!!
* Step 3 🡪 Use ‘**Standard Normal Distribution’** table to find the probability for Z.
  + Table ALWAYS gives probability LESS THAN Z: P(Z < z).

A picture containing text, lamp

Description automatically generated

* + Examples (How to use table)
  + Left probability = TABLE (Directly)

A picture containing text, lamp

Description automatically generated

* + Right probability = 1 - LEFT

* + A picture containing text, lamp

    Description automatically generatedBetween probability = LEFT Z2 – LEFT Z1

**11.5 – Confidence Intervals**

Point Estimates (PE)

* Using a statistic to estimate a parameter
  + Proportions: and Means:

A diagram of a wire with text

Description automatically generated with medium confidence

Margin of error

* C.I. = Point Estimate ± Margin of Error
  + is the distance we extend our guess in both directions to form an interval
* Rule of thumb for margin of error in a survey
  + With 95% confidence, the margin of error, , is approximately for a sample of size
* Interpretation (general structure)
  + A diagram of a diagram

    Description automatically generatedI am C% confident that the true/population parameter + context is between (lower bound) and (upper bound).
* Comparing confidence intervals
  + When comparing confidence intervals to a particular value, or other intervals, we need to look at the ENTIRE interval to see if it is COMPLETELY below or above our comparison.
* Example: Out of 688 randomly selected students, 223 are members of at least one school club.

a) Find the point estimate.

b) Find the lower and upper bounds of a 95% CI using the rule of thumb to calculate the margin of error.

Calendar

Description automatically generated**12.3 – Data Exploration**

Scatterplots:

* **Form**: Linear, curved, or random scatter
* **Direction**: Positive, negative or no association
* **Strength**: Weak, moderate or strong

Timeline

Description automatically generatedCorrelation (*r*):

* Interpreting correlation (LINEAR)
  + Sign = Direction
  + Absolute value |*r*| = Strength

Example

Dataset:

a) Calculate the correlation *r*.

b) Determine if *r* is significant

for α = 0.01.

c) Suppose we have different regression equation where

.

Predict *Y* for *X = 3*:

* Calculate using calculator
  + **LinReg(ax+b) or 2-Var Stats**
  + *L1 = X, L2 = Y*



Regression:

Table

Description automatically generated with medium confidence

* Step 1 🡪 Determine if there is a **significant correlation**

(**linear relationship**).

* + Compare |*r*| and Critical Value (CV)

for *n* (sample size) and significance level α.

* + If |*r*| > CV 🡪 statistically significant.
* Step 2 🡪 Once we have a significant correlation,

we can find the **regression line**.

* + (get results from correlation calculation)

*= slope x + intercept*

Chart, scatter chart

Description automatically generated

* Step 3 🡪 Make **predictions** using the regression line.
  + Just plug in the new *X* value to our equation

and this will give us the predicted *Y.*