Descriptive Statistical Measures

Raw Data	Ungrouped Frequency Distribution	Grouped Frequency Distribution

Measures of the Centre:

Raw Data	Ungrouped Frequency Distribution	Grouped Frequency Distribution
$\mu = \dfrac{\sum x}{N}$ or $\bar{x} = \dfrac{\sum x}{n}$	$\mu, \bar{x} = \dfrac{\sum xf}{\sum f}$	$\mu, \bar{x} = \dfrac{\sum xf}{\sum f}$
Median Position $= \frac{1}{2}(N+1)$ or $\frac{1}{2}(n+1)$	* Median Position $= \frac{1}{2}\left(\sum f + 1\right)$	* Median Position $= \frac{1}{2}\left(\sum f\right)$
Median Value $= x_{\frac{1}{2}(N+1)}$ or $x_{\frac{1}{2}(n+1)}$	* Median Value $= x_{\frac{1}{2}(\sum f+1)}$	* Median Value $= L_i + \dfrac{\left\{\frac{1}{2}\left(\sum f\right) - <Cf_{i-1}\right\}}{f_i} \cdot \Delta x$

Measures of Dispersion:

Raw Data	Ungrouped Frequency Distribution	Grouped Frequency Distribution
a.d. $= \dfrac{\sum \lvert x-\mu \rvert}{N}$ or $\dfrac{\sum \lvert x-\bar{x} \rvert}{n}$		
$\sigma = \sqrt{\dfrac{\sum (x-\mu)^2}{N}}$		
$\sigma = \sqrt{\dfrac{\sum x^2}{N} - \left(\dfrac{\sum x}{N}\right)^2}$	$\sigma = \sqrt{\dfrac{\sum x^2 f}{\sum f} - \left(\dfrac{\sum xf}{\sum f}\right)^2}$	$\sigma = \sqrt{\dfrac{\sum x^2 f}{\sum f} - \left(\dfrac{\sum xf}{\sum f}\right)^2}$
$s = \sqrt{\dfrac{\sum (x-\bar{x})^2}{n-1}}$		
$s = \sqrt{\dfrac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$	$s = \sqrt{\dfrac{\sum x^2 f - \frac{(\sum xf)^2}{\sum f}}{\sum f - 1}}$	$s = \sqrt{\dfrac{\sum x^2 f - \frac{(\sum xf)^2}{\sum f}}{\sum f - 1}}$

$R = x_n - x_1$ $\quad IQR = Q_3 - Q_1$ $\quad IDR = D_9 - D_1$ $\quad$ 10-90 $PR = P_{90} - P_{10}$

Lower Fence $= Q_1 - 1.5(IQR)$ $\quad$ Upper Fence $= Q_3 + 1.5(IQR)$

Standard Scores

$z = \dfrac{x-\mu}{\sigma}$	$z = \dfrac{x-\bar{x}}{s}$
$x = \mu + z\sigma$	$x = \bar{x} + zs$

Miscellaneous Formulae:

$$P = \frac{f}{\sum f} \qquad p = \frac{P}{\Delta x} \qquad 1 - \frac{1}{k^2}$$

---

* Formulae denoted with an asterisk are from optional sections of the course.

Basic Probability Formulae

Counting Rules

$${}_nP_r = \frac{n!}{(n-r)!} \qquad n^r \qquad {}_nC_r = \frac{n!}{r!(n-r)!}$$

Probability Rules

$$P(A) = \frac{n(A)}{n(S)} \qquad 0 \leq P(A) \leq 1 \qquad P(A) + P(\text{-}A) = 1$$

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \qquad P(A \text{ and } B) = P(A) \cdot P(B|A)$$

$$P(A \text{ or } B) = P(A) + P(B) \qquad P(A \text{ and } B) = P(A) \cdot P(B)$$

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

Probability Distributions Formulae

Binomial Probability Distribution Function

$$P(x) = {}_nC_x \cdot \pi^x \cdot (1-\pi)^{n-x}$$

Parameters of Probability Distributions

General Formulae for Discrete Distributions

$$\mu = E(x) = \sum xP \qquad \sigma = \sqrt{\sum x^2P - \left(\sum xP\right)^2}$$

Shortcuts for the Binomial Distribution

$$\mu = E(x) = n\pi \qquad \sigma = \sqrt{n \cdot \pi \cdot (1-\pi)}$$

One Population Inferences Formulae

Standard Error for Single Means and Proportions

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad s_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad \sigma_p = \sqrt{\frac{\pi \cdot (1-\pi)}{n}} \qquad s_p = \sqrt{\frac{p \cdot (1-p)}{n}}$$

*Finite Correction Factor $\Rightarrow$ $^{*}F.C.F. = \sqrt{\dfrac{N-n}{N-1}}$

Maximum Error of Estimate for Single Means And Proportions

$$E = z \cdot s_{\bar{x}} \qquad E = t \cdot s_{\bar{x}} \qquad E = z \cdot s_p$$

Confidence Intervals for Population Means and Proportions

$$P\left(\bar{x} - E < \mu < \bar{x} + E\right) = 1 - \alpha \qquad P\left(p - E < \pi < p + E\right) = 1 - \alpha$$

Sample Sizes for Estimating Means And Proportions

$$n = \left[\frac{z \cdot \sigma}{E}\right]^2 \qquad n = \pi \cdot (1-\pi) \cdot \left[\frac{z}{E}\right]^2$$

Standard Scores for Single Means and Proportions

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \qquad z = \frac{\bar{x} - \mu}{s_{\bar{x}}} \qquad t = \frac{\bar{x} - \mu}{s_{\bar{x}}} \qquad z = \frac{p - \pi}{\sigma_p}$$

Other Formulae

$$_{N}C_{n} \ \text{ or } \ N^n \qquad p = \frac{x}{n} \qquad df = n - 1$$

Bivariate Data Analysis

Linear Regression

$$y_p = a + bx$$

$$a = \frac{\left(\sum x^2\right) \cdot \left(\sum y\right) - \left(\sum x\right) \cdot \left(\sum xy\right)}{n \cdot \left(\sum x^2\right) - \left(\sum x\right)^2} \qquad {}^{*}a = \bar{y} - b\bar{x}$$

$$b = \frac{n \cdot \left(\sum xy\right) - \left(\sum x\right) \cdot \left(\sum y\right)}{n \cdot \left(\sum x^2\right) - \left(\sum x\right)^2} \qquad {}^{*}b = r\left(\frac{s_y}{s_x}\right)$$

Correlation Analysis

$$r = \frac{n \cdot \left(\sum xy\right) - \left(\sum x\right) \cdot \left(\sum y\right)}{\sqrt{n \cdot \left(\sum x^2\right) - \left(\sum x\right)^2} \cdot \sqrt{n \cdot \left(\sum y^2\right) - \left(\sum y\right)^2}} \qquad {}^{*}r = \frac{s_{xy}}{s_x \cdot s_y}$$

Covariance

$${}^{*}s_{xy} = \frac{\sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}}{n-1}$$

Multiple Population Inferences Formulae

Standard Error Formulae for Differences of Means and Proportions

$$^*s_{\bar{d}} = \frac{s_d}{\sqrt{n}} \qquad s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \qquad s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(n_1-1)\cdot s_1^2+(n_2-1)\cdot s_2^2}{n_1+n_2-2}\cdot\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}$$

$$^*s_{p_1-p_2} = \sqrt{\frac{p_1\cdot(1-p_1)}{n_1}+\frac{p_2\cdot(1-p_2)}{n_2}} \qquad ^*s_{p_1-p_2} = \sqrt{p_{\text{Pool}}\cdot(1-p_{\text{Pool}})\cdot\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}$$

Maximum Error of Estimate for Differences of Means and Proportions

$$^*E = t\cdot s_{\bar{d}} \qquad E = z\cdot s_{\bar{x}_1-\bar{x}_2} \qquad E = t\cdot s_{\bar{x}_1-\bar{x}_2} \qquad ^*E = z\cdot s_{p_1-p_2}$$

Confidence Intervals for Differences of Means and Proportions

$$^*P\left(\bar{d}-E<\mu_1-\mu_2<\bar{d}+E\right)=1-\alpha$$

$$P(\,[\bar{x}_1-\bar{x}_2]-E<\mu_1-\mu_2<[\bar{x}_1-\bar{x}_2]+E\,)=1-\alpha$$

$$^*P(\,[p_1-p_2]-E<\pi_1-\pi_2<[p_1-p_2]+E\,)=1-\alpha$$

Standard Scores For Differences Between Means And Proportions†

$$^*t=\frac{\bar{d}-0}{s_{\bar{d}}} \qquad z=\frac{(\bar{x}_1-\bar{x}_2)-0}{s_{\bar{x}_1-\bar{x}_2}} \qquad t=\frac{(\bar{x}_1-\bar{x}_2)-0}{s_{\bar{x}_1-\bar{x}_2}} \qquad ^*z=\frac{(p_1-p_2)-0}{s_{p_1-p_2}}$$

†Here we are assuming hypothesis tests of equality so $\mu_1-\mu_2=0$ and $\pi_1-\pi_2=0$ in these formulae already.

Other Formulae

$$^*d = x_1 - x_2 \qquad df = n_1+n_2-2 \qquad ^*p_{\text{Pool}}=\frac{x_1+x_2}{n_1+n_2}$$

ANOVA Table

Source	df	Sum of Squares	Mean Squares	F statistic
Treatments	$df_T = k-1$	$SST=\sum\frac{T_i^2}{n_i}-CM$	$MST=\frac{SST}{df_T}$	$F=\frac{MST}{MSE}$
Error	$df_E = n-k$	$SSE=\text{Total }SS-SST$	$MSE=\frac{SSE}{df_E}$	
Total	$df_{TOT}=n-1$	$\text{Total }SS=\sum x^2-CM$		$\left(F_{\text{crit}}:\ \begin{array}{l}df_1=df_T\\ df_2=df_E\end{array}\right)$

Here $CM=\frac{\left(\sum x\right)^2}{n}=\frac{\left(\sum T_i\right)^2}{n}$ where T_i is the sum of x for the i^{th} treatment with size n_i.

Requirements for a Complete Solution (F.S.A.R.U.)

Formula State the formula used, including appropriate symbols.

Substitution Substitute the values for your problem. Remember to add any columns to a table that are required for the calculation of those values.

Answer Write your answer including sufficient extra decimals of significance to allow for rounding.

Roundoff Round to the appropriate number of decimal places.

Units Include the appropriate units.

Steps for a Confidence Interval

Step 1) Identify all given information with symbols, preferably on a Venn diagram.

Step 2) Draw a diagram of the sampling distribution.

Step 3) Determine the z-value or t-value.

Step 4) Calculate the standard error.

Step 5) Calculate the maximum error of estimate, E .

Step 6) Make the confidence interval statement.

Steps for a Hypothesis Test

Critical Value Approach

Step 1) Formulate the null and alternative hypotheses

Step 2) State the level of significance.

Step 3) Determine the test statistic.

Step 4) Establish a decision rule.
(draw sampling distribution, find critical value)

Step 5) Evaluate the evidence.
(identify statistics on a Venn diagram, find the calculated value)

Step 6) State your decision.

P-value Approach

Step 1) Formulate the null and alternative hypotheses

Step 2) State the level of significance.

Step 3) Determine the test statistic.

Step 4) Evaluate the evidence.
(identify statistics on a Venn diagram, find the calculated value)

Step 5) Find the P-value.
(draw sampling distribution)

Step 6) State your decision.