Mathematical Notation

Math 113 - Introduction to Applied Statistics

Name : _____

Use Word or WordPerfect to recreate the following documents. Each article is worth 10 points and can be printed and given to the instructor or emailed to the instructor at james@richland.edu. If you use Microsoft Works to create the documents, then you must print it out and give it to the instructor as he can't open those files.

Type your name at the top of each document.

Do not create the watermark $f(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$ on your document.

This is in there so you don't just photocopy the document and give it back to me.

For expressions or equations, you should use the equation editor in Word or WordPerfect. The documents were created using a 14 pt Times New Roman font with standard 1" margins.

For individual symbols (μ, σ, etc), you can insert symbols. In Word, use "Insert / Symbol" and choose the Symbol font. For WordPerfect, use Ctrl-W and choose the Greek set.

The due date for each of these documents is the day after the exam for that chapter. While the material is not due until after the exam, it is recommended that you create it ahead of time because the material will help you review for the exam.

Chapter 2 - Measures of Variation

Attempt #1

The first attempt that people would make is to find the difference between the values and the mean and then add these together. Since this sum, $\sum\left(x-\overline{x}\right)$, is always 0, it is totally meaningless as a measure of spread.

Effort #2

We need some way to make sure the values don't cancel each other out. We can do that by making them positive, either by taking the absolute value or by squaring them. We are going to square them and then add them up. This is called the variation.

VARIATION = Sum of the squared deviations from the mean.

$$Variation = \sum\left(x-\overline{x}\right)^2$$

There is a problem with the variation. It isn't really an "average" spread, just a total spread.

Pass #3

We need to divide the variation to get an average spread.

VARIANCE = average squared deviation from the mean.

$$s^2 = \frac{variation}{df} = \frac{\sum\left(x-\overline{x}\right)^2}{n-1} \qquad \sigma^2 = \frac{variation}{n} = \frac{\sum\left(x-\mu\right)^2}{n}$$

The problem with the variance is that the units are squared. If our original data had units of dollars, then the variation and variance both have units of square dollars. What is a square dollar? Don't know? Exactly! So we need to fix that.

Try #4

We are going to take the square root of the variance to bring the measure of spread back to the original units. This is called the standard deviation.

STANDARD DEVIATION = average deviation from the mean.

$$s = \sqrt{s^2} \qquad \sigma = \sqrt{\sigma^2}$$

Chapter 3 - Probability Distributions

A probability distribution is a list of all the values a random variable can assume with their associated probabilities.

There are two rules for a distribution to be a probability distribution.

1. The probabilities must be between 0 and 1 inclusive.

$$0 \le p(x) \le 1$$

2. The sum of all the probabilities of disjoint events must be 1.

$$\sum p(x) = 1$$

There are formulas for finding the mean, variance, and standard deviation of a probability distribution.

Mean

$$\mu = \sum x \cdot p(x)$$

Variance

$$\sigma^2 = \sum x^2 \cdot p(x) - \mu^2$$

Standard Deviation

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

While it is true that these formulas exist, it is much easier to use the PDIST program that the instructor wrote for the TI82 or TI83 calculators. To use the program, you put the data values x into L_1 and the probabilities $p(x)$ into L_2 and then run the program.

Chapter 5 - Central Limit Theorem

The Central Limit Theorem applies to the sampling distribution of the sample means $\overline{x}$ when samples of size *n* are taken from a population with a mean of μ and a standard deviation of σ.

1. The mean of the sample means is equal to the mean of the population.

$$\mu_{\overline{x}} = \mu$$

2. The variance of the sample means is equal to the variance of the population divided by the sample size.

$$\sigma_{\overline{x}}^2 = \frac{\sigma^2}{n}$$

3. The standard deviation of the sample means (also known as the "standard error of the mean") is the population standard deviation divided by the square root of the sample size.

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

4. The sample means will be normally distributed if the parent population is normally distributed or approximately normally distributed if the sample size is sufficiently large ($n \geq 31$).

For an individual value *x*, use $z = \frac{x - \mu}{\sigma}$

For the mean of a sample $\overline{x}$, use $z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$

Chapter 7 - Hypothesis Testing

All hypothesis testing is done under the assumption that the null hypothesis is true. If the results we get are too unusual to happen by chance alone, then we reject our assumption that the null hypothesis is true.

The null hypothesis H_0 is a statement of no change and always contains the equal sign. Our decision is always based on the null hypothesis and is either to reject the null hypothesis or fail to reject the null hypothesis. If the claim involves the null hypothesis, then we will either have enough evidence to reject the claim or we won't have enough evidence to reject the claim, but we will never accept or support the null hypothesis.

The alternative hypothesis H_1 is a statement of change and never contains the equal sign. If the claim is the alternative hypothesis, then we will either have enough evidence to support the claim or we won't have enough evidence to support the claim, but we won't reject the claim.

The p-value is the probability of getting the results we did if the null hypothesis is true. The level of significance, α, is how unusual we require something to be before saying it's too unusual. We will reject the null hypothesis is the p-value is less than the level of significance and fail to reject the null hypothesis if the p-value is greater than the level of significance.

Mean

With σ known or a large sample size, use $z = \dfrac{\overline{x} - \mu}{\sigma / \sqrt{n}}$

With σ unknown and a small sample size, use $t = \dfrac{\overline{x} - \mu}{s / \sqrt{n}}$ with $df = n - 1$

Proportion

If the expected frequency of each category is at least five, then use $z = \dfrac{\hat{p} - p}{\sqrt{pq / n}}$

Standard Deviation

If the population is essentially normal, then use $\chi^2 = \dfrac{df \bullet s^2}{\sigma^2}$ with $df = n - 1$

Chapter 9 - Correlation and Regression

Correlation is a measure of the strength of a relationship. Regression describes that relationship.

Pearson's Linear Correlation Coefficient ρ (or *r* for a sample) describes the strength of a linear relationship. Here are some properties of *r* (or ρ).

1. *r* is always between -1 and 1. $-1 \le r \le 1$
2. *r* only measures the strength of a linear relationship.
3. *r* is unchanged if either variable is rescaled.
4. *r* is unchanged if the variables are switched.

The formula for *r* is $r = \dfrac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$, but it is much easier to let a computer or calculator find it for us than using the formula.

The regression equation, also known as the best fit line, can be written as $\hat{y} = ax + b$ where $a = \dfrac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$ and $b = \bar{y} - a\bar{x}$. It is much easier to let technology find these values for us, though.

The regression equation always passes through the centroid $(\bar{x}, \bar{y})$ of the data and so if there is no significant linear correlation, then the best equation to use is that of a horizontal line passing through the centroid, $\hat{y} = \bar{y}$.

The coefficient of determination is the percent of the variation that can be explained by the regression equation and can be found by using the formula $r^2 = \dfrac{\text{explained variation}}{\text{total variation}}$ or simply by squaring the correlation coefficient.