

# Mathematical Notation

## Math 113 - Introduction to Applied Statistics

Use Word or WordPerfect to recreate the following documents. Each article is worth 10 points and should be 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. Include the title as part of what you type. The lines around the title aren't that important, but if you will type ----- at the beginning of a line and hit enter, both Word and WordPerfect will draw a line across the page for you.

For expressions or equations, you should use the equation editor in Word or WordPerfect. The instructor used WordPerfect and a 14 pt Times New Roman font with 0.75" margins, so they may not look exactly the same as your document. The equations were created using 14 pt font and appear a little larger than normal text. This is a quirk of WordPerfect, but it allows you to see what was done with the equation editor versus the word processor.

For individual symbols ( $\mu$ ,  $\sigma$ , etc) within the text of a sentence, you can insert symbols. In Word, use "Insert / Symbol" and choose the Symbol font. For WordPerfect, use Ctrl-W and choose the Greek set. However, it's often easier to just use the equation editor as expressions are usually more complex than just a single symbol. If there is an equation, put both sides of the equation into the same equation editor box instead of creating two objects.

There are instructions on how to use the equation editor in a separate document or on the website. Be sure to read through the help it provides. There are some examples at the end that walk students through the more difficult problems. You will want to read the handout on using the equation editor if you have not used this software before.

If you fail to type your name on the document, you will lose 1 point. Don't type [the hints or the reminders](#) at the bottom of each page.

These notations are due before the beginning of class on the day of the exam for that material. For example, notation 1 is due on the day of exam 1. Late work will be accepted but will lose 20% of its value per class period.

---

## Notation 1: Chapters 1 - 3

---

The spread of a distribution is described by the standard deviation, and finding it by hand is a four step process.

1. The sample mean is found by adding up all the values and dividing by the number

of values.  $\bar{x} = \frac{\sum x}{n}$

2. The variation is also known as the sum of the squares (SS) because you find how far each value deviates from the mean, square that deviation, and then sum those

squares.  $Variation = \sum (x - \bar{x})^2$

3. The sample variance is also known as a mean square (MS) and is found by dividing the variation by the degrees of freedom (which is  $n - 1$  in this case).

$$Variance = s^2 = \frac{Variation}{df}$$

4. The standard deviation is found by taking the square root of the variance.

$$s = \sqrt{Variance}$$

Consider the five numbers 3, 8, 2, 4, and 7. The sum of the original values is 24, so divide that by 5 to get the mean of  $\bar{x} = 4.8$ .

|                   |      |       |      |      |      | Total |
|-------------------|------|-------|------|------|------|-------|
| $x$               | 3    | 8     | 2    | 4    | 7    | 24    |
| $x - \bar{x}$     | -1.8 | 3.2   | -2.8 | -0.8 | 2.2  | 0     |
| $(x - \bar{x})^2$ | 3.24 | 10.24 | 7.84 | 0.64 | 4.84 | 26.8  |

In the table above, the 26.8 is the variation. Divide that by  $5-1=4$  to get the variance of  $s^2 = 6.7$ . The standard deviation is  $s = \sqrt{6.7} \approx 2.59$ .

As you can see, that's pretty involved, so most of the time, we'll just let technology find the standard deviation for us.

To standardize a variable, we take the deviation from the mean and divide it by the standard deviation. This is called a z-score and is found using the formula  $z = \frac{x - \bar{x}}{s}$ .

The z-score is a ruler for how many standard deviations a value is from the mean. Standardizing doesn't change the shape of a distribution, but it makes the mean 0 and the standard deviation 1.

The probability of an event is its long-run relative probability. The law of large numbers says that the long-run relative frequency of an event will get closer and closer to the true relative frequency as the number of trials increases. However, there is no law of averages that applies to the short-term.

All probabilities must be between 0 and 1 inclusive,  $0 \leq P(A) \leq 1$ . If an event is impossible, then the probability of it happening is 0. If an event must happen, then the probability of it happening is 1. The sum of the probabilities of all the different outcomes is one.

If all the outcomes are equally likely, then the probability of something happening is the number of ways it can occur divided by the total number of outcomes.

The complement of an event is everything that is not that event. The probability that something won't happen is one minus the probability that it will happen,  
 $P(A') = 1 - P(A)$ .

If two events are mutually exclusive (disjoint), then the probability of one or the other occurring is the sum of their probabilities. If two events are independent (not related), then the probability of both occurring is the product of their probabilities.

Tree diagrams are useful for finding probabilities of compound events. To find the probability of reaching the end of a branch, you multiply the probabilities on all the branches getting to that point. At each point in a tree diagram, the sum of the probabilities of all the branches from a single point must be one.

---

## Notation 2: Chapters 4 - 6

---

The mean of a discrete probability distribution is also called its expected value and is

$\mu = E(x) = \sum xp(x)$ . The variance of a probability distribution is

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

A binomial experiment is a fixed number of independent trials each having exactly two possible outcomes. The mean of a binomial distribution is  $\mu = np$  and the standard

deviation is  $\sigma = \sqrt{npq}$ .

The normal distribution is unimodal, symmetric, and bell-shaped. The empirical rule states that all normal distributions have approximately 68% of the data within one standard deviation of the mean, 95% of the data within two standard deviations of the mean, and 99.7% of the data within three standard deviations of the mean.

Any normal model can be converted into a standard normal distribution by standardizing

the variable by subtracting the mean and dividing by the standard deviation,  $z = \frac{x - \mu}{\sigma}$ .

Provided that the sampled values are independent and the sample size is large enough, the sampling distribution of the sample proportions,  $\hat{p}$ , is modeled by a normal model with

the  $Mean(\hat{p}) = p$  and the  $SD(\hat{p}) = \sqrt{\frac{pq}{n}}$ . If we don't know the population

proportion  $p$ , then we'll use the standard error of the proportion,  $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$ .

Provided that the sampled values are independent and the sample size is large enough, the sampling distribution of the sample means,  $\bar{x}$ , is modeled by a normal model with the

$Mean(\bar{x}) = \mu$  and the  $SD(\bar{x}) = \frac{\sigma}{\sqrt{n}}$ . We usually don't know the population

standard deviation  $\sigma$ , so we use the standard error of the mean,  $SE(\bar{x}) = \frac{s}{\sqrt{n}}$ , with the

Student's t distribution.

The Student's t distribution is very similar to the standard normal distribution except the standard deviation is greater than one. Graphically, this means that the curve is flatter in the middle and wider in the tails than the normal curve. There are actually many t distributions, one for each degree of freedom, but as the sample size increases, the t distributions approach the standard normal distribution.

The critical value is based on the model being used and the level of confidence. The center of the confidence interval for the population parameter is the sample statistic ( $\hat{p}$  or  $\bar{x}$ ). The distance from the center of the interval to either endpoint is called the margin of error (ME) or maximum error of the estimate. The margin of error is the critical value times the standard error,  $ME = CV \times SE$ .

---

### **Notation 3: Chapters 7 - 8**

---

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 from the normal or assumed condition and always contains the equal sign. Our decision is always based on the null hypothesis and is either to reject or retain the null hypothesis. If the claim involves the null hypothesis, then we will use the word "reject" in our conclusion. We will never accept or support the null hypothesis.

The alternative hypothesis,  $H_1$ , is a statement of change from the normal or assumed condition and never contains the equal sign. The alternative hypothesis is used to determine whether the test is a left tail, right tail, or two tail test. If the claim is the alternative hypothesis, then we will use the word "support" in our conclusion.

The critical value is a pre-determined value based on the model, not on the sample data. It separates the critical region, where the null hypothesis is rejected, from the non-critical region, where the null hypothesis is retained. The test statistic is a value that is based on the sample data and the decision to reject or retain the null hypothesis is based on which region the test statistic falls into.

The p-value is the probability of getting the results we did if the null hypothesis is true. The level of significance,  $\alpha$ , is how unusual we require something to be before saying it's too unusual to happen by chance alone. We will reject the null hypothesis if the p-value

is less than the level of significance and retain the null hypothesis if the p-value is greater than the level of significance.

Besides looking at the test statistic to see whether or not it lies in the critical region, you can also look at the confidence interval to see whether or not it contains the hypothesized or claimed value. Since the confidence intervals are the believable values, we'll retain the null hypothesis if the claimed value falls in the confidence interval.

When working with two independent samples, we need to know how variables behave when we combine them. The mean of a difference is the difference of the means  $Mean(x - y) = Mean(x) - Mean(y)$ , but the variance of a difference is the sum of the variances  $Var(x - y) = Var(x) + Var(y)$ .

When working with two independent means, you can use two formulas, one with a pooled variance and one without. Most of the time, it is safer and easier to assume the variances are not equal and not pool them together. Sometimes it makes sense to assume that the variances are equal and we'll make that assumption when we work Analysis of Variance problems in chapter 10.

If you have paired data, then you create a new variable  $d$  that is the difference between the two observations and then work it out as a test about a single population mean.

---

## **Notation 4: Chapters 9 - 10**

---

The correlation coefficient is a measure of how well the line fits the data. The correlation coefficient,  $r$ , is always between -1 and +1. Correlations near zero correspond to weak or no linear correlation. Changing the order of the  $x$  and  $y$  variables won't change the value of  $r$ . Changing the scale on either variable won't change the value for  $r$  because it is based on the standardized score. There are no units on  $r$ . Correlation is sensitive to outliers. Do not correlate categorical variables, check the scatter plot to make sure the association is linear, and beware of outliers. Correlation does not imply causation.

When looking at a scatter plot, there are three things you should look for. Form: is the plot linear, quadratic, or some other shape? Scatter: how spread out is the data? and Direction: does the pattern have a positive direction or a negative one?

The best fit line will always pass through the centroid of the data, which is the point  $(\bar{x}, \bar{y})$ , and the equation of the line is  $\hat{y} = b_0 + b_1x$  where  $b_0$  and  $b_1$  are the y-intercept

and slope of the line. The slope is related to the correlation coefficient using the equation  $b_1 = r \frac{s_y}{s_x}$ . If there is no significant correlation, then the best fit line is horizontal and the estimated value is always the mean,  $\hat{y} = \bar{y}$ .

The coefficient of determination,  $r^2$ , is the percent of the variation that can be explained by the regression equation. The higher the coefficient of determination, the better the model, but there is no magic number for how large it should be to consider the model good.

For multiple regression, there is one response variable and several predictor variables. One should look at the adjusted- $R^2$ , rather than the  $R^2$ , when determining the best model. The adjusted- $R^2$  takes into account the sample size and the number of independent variables. The  $R^2$  and adjusted- $R^2$  have similar formulas, with the  $R^2$  using the variations (SS) while the adjusted- $R^2$  uses the variances (MS).

$$R^2 = \frac{SS(\text{total}) - SS(\text{residual})}{SS(\text{total})} \quad \text{Adj} - R^2 = \frac{MS(\text{total}) - MS(\text{residual})}{MS(\text{total})}$$

The  $\chi^2$  goodness of fit test checks the hypothesis of whether the claimed proportions are correct. It does this by comparing the observed frequency of categories to their expected frequencies and seeing how close they are. If the observations are close to what is expected, then the test statistic is small and we would retain the null hypothesis while large differences would mean that the claim is wrong and we would reject the null hypothesis. Therefore, the goodness of fit test is therefore always a right tail test. The degrees of freedom is one less than the number of categories and the test statistic is

$$\chi^2 = \sum \frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}}$$

The contingency table can be used to check for independence between two categorical variables. It uses the same formula for the test statistic as the goodness of fit test, but the degrees of freedom is  $df = df_{\text{row}} \times df_{\text{col}}$ . The expected frequency for each cell is Row Total  $\times$  Column Total  $\div$  Grand Total.

The one-way ANOVA compares more than two means to see if they're equal. The null hypothesis is that the means are equal and the alternative is that at least one of the means

is different. The "one-way" part is because the data is categorized in exactly one way, much like a goodness of fit test.

Here is what a typical one-way ANOVA table looks like.

| Source           | SS   | df | MS     | F    | p     |
|------------------|------|----|--------|------|-------|
| Between (Factor) | 1820 | 4  | 455    | 3.50 | 0.033 |
| Within (Error)   | 1950 | 15 | 130    |      |       |
| Total            | 3770 | 19 | 198.42 |      |       |

The values in the SS (Sum of Squares) column are variations and the values in the MS (Mean Square) column are sample variances. To find the MS, you divide the SS by the df. The F test statistic is the ratio of two sample variances and is found by dividing the MS for the row by the MS(Error).

The two-way ANOVA actually has three tests rolled into one. The data is categorized two ways instead of one, much like a test for independence, and two one-way ANOVAs are performed, one for each way of classification. The other test is to see if there is any interaction between the two classification systems.