

Math 214

Lab - Review Descriptive Stats Stuff

Project Laboratory Review #1

Due: NOT due - use to review for Exam1

Let's review some of what we learned about descriptive statistics using R to provide the data.

Some Sample Test Scores - Mean, Median, Mode

Lets do some basic tasks to organize and describe some data. First, we produce a set of 400, normally distributed test scores using R:

```
> scores = rnorm(400,75,8) %% Mean = 75, sd = 8, 400 scores.
```

Now, take a sample of 10 scores from this set to use as your data.

```
> myscores = sample(scores,10) %% Randomly pick 10 of the 400 scores.
> myscores
[1] 72.94076 85.99613 90.14628 53.65725 70.90921 69.14693 78.84987 64.64034
[9] 77.29426 81.34520
```

To make life easier, lets round everything off to whole numbers:

```
> format(myscores,digits = 2)
[1] "73" "86" "90" "54" "71" "69" "79" "65" "77" "81"
```

TO DO:

1. **Central Tendency:** Find the mean, median and mode of your data by hand. (Check it with R.)
2. Subtract 20 from the LOWEST test score and recompute the mean and median. Which one changes the most? Explain whether the mean or the median is a *resistant measure of central tendency*.
3. **Measures of Spread** Find, by hand, the following measures of spread for your data:
 - (a) The range
 - (b) The standard deviation
 - (c) The First and Third Quartiles and the IQR.
4. **Graphical Displays** For your data, construct the following pictures:
 - (a) A dot-plot.
 - (b) A histogram with 10 bins (1-10,11-20,etc,91-100).
 - (c) A box plot.

1. **Z-Scores** Find the z-scores for your 10 test scores.

- (a) What percentage of scores are below 0?
- (b) What percentage of scores are above 1?
- (c) What percentage of scores are below 2?
- (d) How well does your data compare to the 'Empirical Rule'?

2. Use R to find the z-scores for the full set of 400 scores and look at the probability distribution.

```
> z = (scores - mean(scores))/sd(scores)
> hist(z,prob=T)
```

- (a) What percentage of scores are below 0?
- (b) What percentage of scores are above 1?
- (c) What percentage of scores are below 2?
- (d) How well does the full data set compare to the 'Empirical Rule'?

3. Linear Regression: Do problem 3.53 in Kitchens concerning the relationship between repair costs and age of cash registers. The data is:

```
> age
[1] 6 7 1 3 6 2 5 4 3
> cost
[1] 92 181 23 40 126 35 86 72 51
```

- (a) Calculate the Spearman and Pearson correlation coefficients.
- (b) Calculate the best fit least-squares regression line with age as the predictor variable.
- (c) Find the residual at (3,40)
- (d) Redo the calculations WITHOUT the data for 7 year old cash registers.
- (e) What happens to the line and to the error at (3,40)? Why?