Econometrics Review for Midterm exam

1. True / false / uncertain? Explain your answer since the grade will depend upon the explanation. “If there are *N*  independent observations of a random variable with standard deviation then the standard deviation of the sample mean is and the standard deviation of the sample median is 1.25•. Thus as long as a random variable has a well-defined then the sample mean always has a smaller standard deviation than the sample median.”

-The standard error is the sample’s standard deviation divided by √n. It is used to estimate the standard deviation of the sample mean based on the population mean.

-The standard error of sample median should be : standard error of mean \* 1.25

It is thus less efficient and more subject to sampling fluctuations. It is usually to estimate a data that are not drawn from a normally distributed population and rarely used.

It is possible for a sample mean to have higher standard deviation than sample median:

A, sample size, a small sample evident the difference between the data. Therefore it comes to working with samples calculated according to the standard deviation.

B, Outliers, The mean is sensitive to these values and increase the standard deviation.

C, The distribution of the data does not interfere definitively on the standard deviation

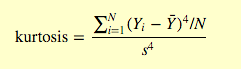
D, if considered analysis of the standard deviation, kurtosis and skewness always be analyzed.

one probem of the mean is their sensitivity to outliers therefore often not enough to analyze the standard deviation. Then, with that information is not enough to make a good decision, analyze the skewness and kurtosis. This allows you to know the presence of outliers.

1. What is the kurtosis of a random variable? How does the kurtosis of random variable X+Y related to the kurtosis of random variables X and Y? What happens if both X and Y are normal? If neither X nor Y is normal?

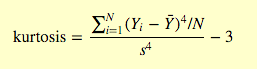
The kurtosis is the fourth [standardized moment](https://en.wikipedia.org/wiki/Standardized_moment), defined as

For univariate data *Y*1, *Y*2, ..., *YN*, the formula for kurtosis is:



is the mean, ***s*** is the standard deviation, and *N* is the number of data points. Note that in computing the kurtosis, the standard deviation is computed using *N* in the denominator rather than *N* - 1.

The kurtosis for a [standard normal distribution](http://www.itl.nist.gov/div898/handbook/eda/section3/eda3661.htm) is three. For this reason, some sources use the following definition of kurtosis (often referred to as "excess kurtosis"):



Var(X+Y)=Var(X)+Var(Y) always hold

Kurtosis indicated how the peak and tails of a distribution differ from the normal distribution. Use kurtosis to help you initially understand general characteristics about the distribution of you data.

Positive kurtosis indicated that distribution has heavier tails and a sharper peak than the normal distribution.

Negative kurtosis value indicates that the distribution has lighter tails and a flatter peak than the normal distribution

Data follow a normal distribution perfectly have a kurtosis that = 0. Normally distributed data establishes the baseline for kurtosis. Sample kurtosis that significantly deviates form 0 may indicate that the data is not normally distributed.

A statistical measure used to describe the distribution of observed data around the mean. Used generally in the statistical field, kurtosis describes trends in charts. A high kurtosis portrays a chart with fat tails and a low, even distribution, whereas a low kurtosis portrays a chart with skinny tails and a distribution concentrated toward the mean.

1. Suppose you know that a random variable has a normal distribution. If you wish to minimize the variance of the estimate and you can use either a sample mean or a sample median, which should you use? What if the distribution is Cauchy? What if you know only that the random variable is one of these distributions? How do your answers to all the questions change if you can select a 20% trimmed mean?

In statistics, **estimation** refers to the process by which one makes inferences about a population, based on information obtained from a sample.

An estimate of a population parameter may be expressed in two ways:

* **Point estimate**. A point estimate of a population parameter is a single value of a statistic. For example, the sample mean x is a point estimate of the population mean μ. Similarly, the sample proportion *p* is a point estimate of the population proportion *P*.
* **Interval estimate**. An interval estimate is defined by two numbers, between which a population parameter is said to lie. For example, *a* < x < *b* is an interval estimate of the population mean μ. It indicates that the population mean is greater than *a* but less than *b*

Two common ways to estimate the center of a set of data are the sample mean and the sample median. The sample mean is sometimes more efficient, but the sample median is always more robust.

When the data come from distributions with thick tails, the sample median is more efficient. When the data come from distributions with a thin tail, like the normal, the sample mean is more efficient.

The sample mean is just the average of the sample values. The median is the middle value when the data are sorted.

Often efficiency and robustness are in tension and you have to decide how much efficiency you’re willing to trade off for how much robustness. ARE gives you a way of measuring the loss in efficiency if you’re right about the distribution of the data but choose a more robust, more cautious estimator. Of course if you’re significantly wrong about the distribution of the data (and often you are!) then you’re better off with the more robust estimator.

Often efficiency and robustness are in tension and you have to decide how much efficiency you’re willing to trade off for how much robustness. ARE gives you a way of measuring the loss in efficiency if you’re right about the distribution of the data but choose a more robust, more cautious estimator. Of course if you’re significantly wrong about the distribution of the data (and often you are!) then you’re better off with the more robust estimator.

-Table, delimma it is that if we don’t know how the data is distributed, we can choose trim which is 2nd best

mean is best for symmetric distributions without outliers

median is useful for skewed distributions or data with outliers

         Mean            Median        Trimmed estimator

Cauchy Horrible           Best                2nd best

Normal                     Best                Bad 2nd best

if we are in a scenario that whether the distribution is cauchy  or normal,

`when the variance is big then suddenly jump into very small number, we assume it is not normally distributed(maybe cauchy but we are not certain about that)

`when the variance is basically the same and small, we say it is probably normally distributed

`when under cauchy, the standard deviation of mean might be over 50 and variance is very large etc, 3000

1. If all you know about the distribution is that is a Tukey (or a mixture of normals) what estimator would you pick if you could select from any one of the trimmed means. Why?
2. Describe how you can use the variance of the empirical distribution to select which estimator to use for each set of data. How does this compare with precommiting to the use of a particular estimator?
3. Prove that the ordinary least squares regression forces the mean of the residuals to zero.

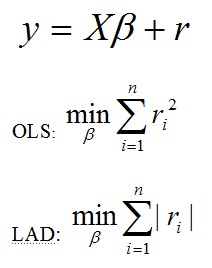
A statistical technique to determine the [line of best fit](http://www.investopedia.com/terms/l/line-of-best-fit.asp) for a model. The [least squares](http://www.investopedia.com/terms/l/least-squares.asp) method is specified by an equation with certain parameters to observed data. This method is extensively used in [regression](http://www.investopedia.com/terms/r/regression.asp) analysis and estimation.

Prove:

1. Prove that the least absolute deviations regression forces the median of the residuals to zero.

Least absolute deviation (LAD) (also known as least absolute residuals, least absolute errors and least absolute value) is an alternative method to the conventional ordinary least squares (OLS), for building [regression](http://www.statisticalconsultants.co.nz/blog/regression.html) models.  Instead of estimating the coefficients that minimise the sum of squared residuals, LAD estimates the coefficients that minimises the sum of the absolute residuals.

Compared to OLS, LAD has the advantage of being resistant to outliers and robust to departures from the normality assumption.    
    
The disadvantages of LAD include being more computationally expensive than OLS and having the possibility of more than one solution. The reason for this is that unlike OLS (which has an analytical solution derived from matrix calculus), the LAD coefficient estimates are found iteratively, usually with the OLS estimates as the initial values.



Prove:

1. Describe the “empirical distribution” in regression context. How might we resample to compute bootstrapped standard errors?

empirical distribution is based on your observation of out comes, it is based on real data. on the other hand theoretical is base on your theory regarding the distribution and the parameters.

The simplest graphical model for relating a response variable y to a single independent variable x is a straight line

We then show how to judge whether a relationship exists between y and x, and how to use the model either to estimate E(y), the mean value of y, or to predict a future value of y for a given value of x. The totality of these methods is called a **simple linear regression analysis**.

The straight-line model is hypothesized to relate sales revenue y to advertising expenditure x. That is,

y = β0 + β1x + ε

1. Describe a t-test for a hypothesized regression coefficient and the decision to reject a hypothesis.

This lesson describes how to conduct a hypothesis test to determine whether there is a significant linear relationship between an independent variable X and a dependent variable Y. The test focuses on the [slope](http://stattrek.com/Help/Glossary.aspx?Target=Slope) of the [regression](http://stattrek.com/Help/Glossary.aspx?Target=Regression) line

Y = Β0 + Β1X, where Β0 is a constant, Β1 is the slope (also called the regression coefficient), X is the value of the independent variable, and Y is the value of the dependent variable.

If we find that the slope of the regression line is significantly different from zero, we will conclude that there is a significant relationship between the independent and dependent variables.

The approach described in this lesson is valid whenever the standard requirements for simple linear regression are met.

The dependent variable Y has a linear relationship to the independent variable X.

For each value of X, the probability distribution of Y has the same standard deviation σ.

For any given value of X,

The Y values are independent.

The Y values are roughly normally distributed (i.e., [symmetric](http://stattrek.com/Help/Glossary.aspx?Target=Symmetry) and [unimodal](http://stattrek.com/Help/Glossary.aspx?Target=Unimodal%20distribution)). A little[skewness](http://stattrek.com/Help/Glossary.aspx?Target=Skewness) is ok if the sample size is large.

Previously, we described [how to verify that regression requirements are met](http://stattrek.com/AP-Statistics-1/Regression.aspx#ReqressionPrerequisites).

The test procedure consists of four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results.

If there is a significant linear relationship between the independent variable X and the dependent variable Y, the slope will not equal zero.

H0: Β1 = 0   
Ha: Β1 ≠ 0

The [null hypothesis](http://stattrek.com/Help/Glossary.aspx?Target=Null%20hypothesis) states that the slope is equal to zero, and the alternative hypothesis states that the slope is not equal to zero.

Using sample data, find the [standard error](http://stattrek.com/Help/Glossary.aspx?Target=Standard%20error) of the slope, the slope of the regression line, the degrees of freedom, the test statistic, and the P-value associated with the test statistic. The approach described in this section is illustrated in the sample problem at the end of this lesson.

* Standard error. Many statistical software packages and some graphing calculators provide the[standard error](http://stattrek.com/Help/Glossary.aspx?Target=Standard%20error) of the slope as a regression analysis output.
* Slope. Like the standard error, the slope of the regression line will be provided by most statistics software packages. In the hypothetical output above, the slope is equal to 35.
* Degrees of freedom. For simple linear regression (one independent and one dependent variable), the [degrees of freedom](http://stattrek.com/Help/Glossary.aspx?Target=Degrees%20of%20freedom) (DF) is equal to:

DF = n - 2

where n is the number of observations in the sample.

* Test statistic. The test statistic is a t-score (t) defined by the following equation.

t = b1 / SE

where b1 is the slope of the sample regression line, and SE is the standard error of the slope.

* P-value. The P-value is the probability of observing a sample statistic as extreme as the test statistic. Since the test statistic is a t-score, use the [t Distribution Calculator](http://stattrek.com/Tables/T.aspx) to assess the probability associated with the test statistic. Use the degrees of freedom computed above.

If the sample findings are unlikely, given the null hypothesis, the researcher rejects the null hypothesis. Typically, this involves comparing the P-value to the [significance level](http://stattrek.com/Help/Glossary.aspx?Target=Significance%20level), and rejecting the null hypothesis when the P-value is less than the significance level. Fail to reject the null hypothesis if the P-value is larger than the significance level.

1. On the basis of the experiments you’ve done, how does the efficiency of a regression predict the power of the t-test?

From the output we see that we can clearly reject the null hypothesis of no linear relationship between height and head circumference (p=9.59e-05).

if x contributes no information for the prediction of y? The implication is that the mean of y (i.e., the deterministic part of the model E(y) = β0 + β1x), does not change as x changes. Regardless of the value of x, you always predict the same value of y. In the straight-line model, this means that the true slope, β1, is equal to 0 (see Figure 3.8.). Therefore, to test the null hypothesis that x contributes no information for the prediction of y against the alternative hypothesis that these variables are linearly related with a slope differing from 0, we test

If the data support the alternative hypothesis, we conclude that x does contribute information for the prediction of y using the straight-line model [although the true relationship between E(y) and x could be more complex than a straight line]. Thus, to some extent, this is a test of the utility of the hypothesized model.

1. When the underling errors in the regression problem are independent normal, show that ordinary least squares is the maximum likelihood estimator?

Prove:

1. Continuing #11. What if the errors are independent Tukey?

Prove: