

What Else Is There?

p-VALUE STATISTICS

- On most of the statistical tests we have done, we compared the table value of a statistic with a calculated value.
 - When we did calculations, we generally ignored everything except the value of the statistic we were testing.
- p-value statistics work “backwards”, but allow you to ignore the table
 - The p-value is the **actual** probability you will say a result is significant when it really just happened by chance.
 - In a p-value test you compare the p-value with the level of significance (α). If $p < \alpha$, the test is significant.
 - To use p-value statistics, you must correctly complete the directional (\neq , $<$, $>$) line in the test input menu.

HIGH-POWER TESTS

- β (beta) is basically the opposite of α . It's the probability you will **overlook** a significant result—think it **isn't** significant when it really is.
- POWER means the ability of a test to control this kind of error.
- A high-power test is very unlikely to overlook significant results.

ANALYSIS OF VARIANCE (ANOVA)

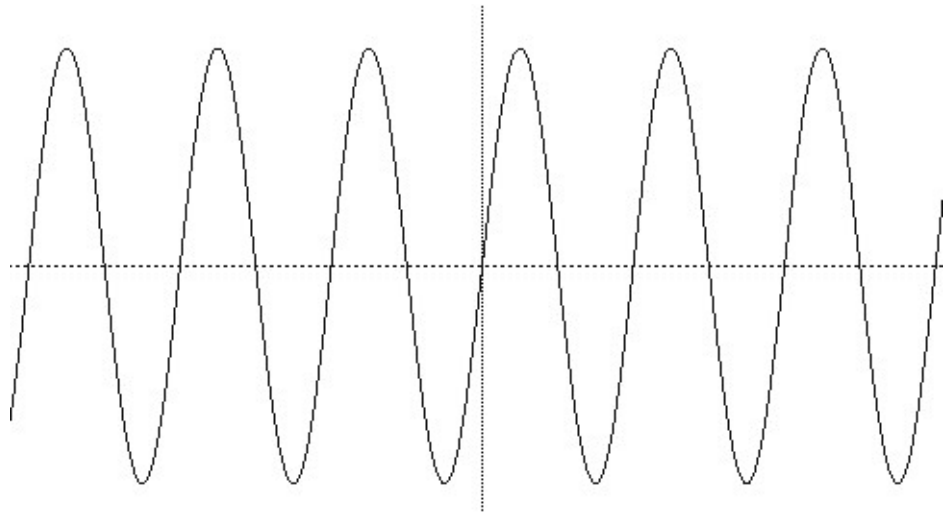
- Also called an F-test
- ANOVA is a way of comparing the mean and/or standard deviation of more than two samples at the same time.
- With what we have learned, the only way we could do this is to match together every possible pair of samples and run t and χ^2 tests on each pair.
 - ➔ For 3 samples, that would be 12 tests
 - ➔ For 5 samples, it would be 240 tests.
- ANOVA essentially runs all those tests at the same time.

SPEARMAN'S “r” TEST

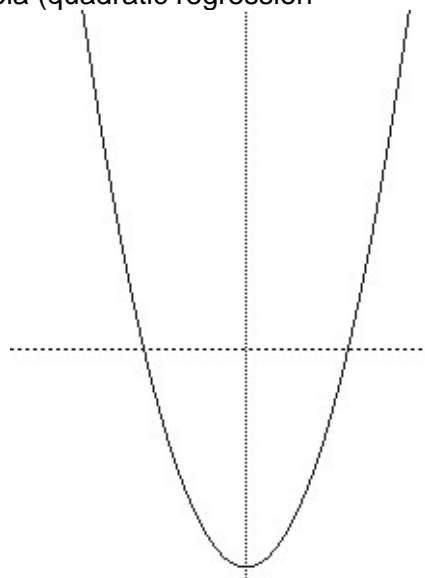
- Used to compare **ranked** data
- Are two sets of rankings similar to or different from each other?
- Examples:
 - Did two judges in the Miss America contest rank the finalists in essentially the same order?
 - Do coaches and sportscasters rank the top teams differently?
- The key thing is **ranking** (1st, 2nd, 3rd, etc.)

NON-LINEAR REGRESSION

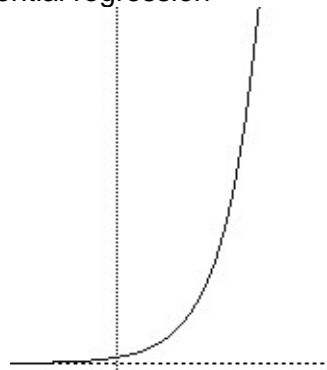
- Looks at patterns that aren't lines.
- Examples are
 - sine wave (sinusoidal regression)



- parabola (quadratic regression)



- exponential regression



MULTIPLE REGRESSION ANALYSIS

- When we discussed linear regression (plugging values of “x” into an equation to predict “y”), we only considered two possible variables.
- However, almost every real-life problem involves more than two variables.
- Multiple regression uses **many** predictors to estimate an outcome.

Example: When Mr. Burrow was in graduate school, one of his professors (Dr. Charles Davidson of the University of Southern Mississippi) was doing exhaustive research on the factors that predict success in college.

He found thirty-seven factors that all had some independent effect on college GPA.

A few of these are:

- family income
- age of student when entering college
- state competency test scores (the equivalent of ITEDs)
- number of high school activities the student participated in
- per-pupil budget of the student's school district
- marital status of parents
- number of brothers and sisters
- academic load (how many semester hours the student takes)
- size of community where the student grew up (in Mississippi, students from larger communities did better than those from small towns or rural areas)

After more than a year of work, he devised a formula that involved the twelve most significant of his 37 predictors. This formula could be used to predict an incoming student's college GPA.

CALCULUS-BASED STATISTICAL THEORY

- Advanced statistics courses typically include a fair amount of applied calculus; some are pretty much all based on theoretical calculus.
- Derivatives (such as $f'(x)$ or $\frac{dy}{dx}$) look at how statistics **change** over time.
- Much of statistics is based on the normal curve, which can be found using definite integrals (another calculus topic). The tables we learned for z-scores and their associated areas are all generated through calculus.