



BIOMETRICS INFORMATION

(You're 95% likely to need this information)

PAMPHLET NO. # 30

DATE: March 5, 1991

SUBJECT: Interpretation of probability p-values

Most statistical computer programs, including SAS, calculate and report probability or p-values in their output. These values can then be used to make conclusions about associated hypotheses. This pamphlet will discuss the use and interpretation of these p-values.

For example, a fertilizer supplier claims that a new type of fertilizer can increase the average growth rate of a certain type of tree. To test this claim, two random samples of these trees are selected and treated with the "new" fertilizer and the standard fertilizer respectively. The objective is to determine if the trees treated with the new fertilizer have a higher growth rate than the trees treated with the standard fertilizer.

In hypothesis testing, two contradictory hypotheses are under consideration. The null hypothesis H_0 states an equality between population parameters, for example, trees treated with the new fertilizer have the same growth rate as trees treated with the standard fertilizer. The alternate hypothesis H_a states a difference between parameters and is usually what the experimenter hopes to verify, for example, trees treated with the new fertilizer have a better growth rate than trees treated with the standard fertilizer. The null hypothesis will be rejected if the sample evidence is more consistent with the alternative hypothesis.

When drawing a conclusion in hypothesis testing, one is faced with the possibility of making two types of errors. A Type I error consists of rejecting H_0 when it is true and a Type II error involves not rejecting H_0 when it is false. The probability of making a type I error is traditionally denoted by α , and the probability of making a type II error by β . Often the value α is specified to indicate the amount of Type I error one is willing to tolerate. This α value is referred to as the **significance level** of the test. Typically acceptable α -levels are 0.05 and 0.01. If, for example, one concludes that the growth rate of trees treated with the new fertilizer is significantly different from those treated with the standard fertilizer at $\alpha = 0.05$, then there is a 5% chance that the conclusion is wrong.

While reporting the result of an hypothesis test, stating that H_0 is rejected at some α level is not sufficient as it does not provide information on the weight of evidence against it. Also, it dictates the α -level that others must use. A more appropriate approach is to report the **p-value**: the smallest level at which the data is significant. H_0 is rejected at the α -level only if the p-value is less than α . Suppose the p-value of the above fertilizer test is 0.02, that is, if one were to reject H_0 based on the data, then there is a 2% chance that the decision is incorrect. At $\alpha = 0.05$, the experimenter would reject H_0 as the Type I error (2%) in the data is within the allowed amount of 5% specified. Conversely, at $\alpha = 0.01$, the experimenter would retain H_0 as the

Type I error in the data exceeds what he/she is willing to tolerate. That is, the test is significant at 5% level but not significant at 1%. In fact, p-value measures the amount of statistical evidence against the null hypothesis in favor of the alternative hypothesis: the smaller the p-value the stronger the evidence against the null hypothesis. When drawing a conclusion, the weight of evidence indicated by the computed p-value should be reported. This weight of evidence may be indicated in words according to the following table:

<u>p-value = p</u>	<u>Weight of evidence against the null hypothesis</u>
$p \leq 0.01$	very strong
$0.01 < p \leq 0.05$	strong
$0.05 < p \leq 0.10$	moderate
$0.10 < p$	little or no

Example 1: Suppose two different fertilizers (1 and 2) are available and one is interested in testing if there is a difference in performance between them. Two samples of trees are randomly selected and each is treated with one of the two fertilizers and the growth rate (Y) is measured. This test can be carried out using PROC ANOVA or PROC GLM with class variable A indicating the type of fertilizer (1 or 2) applied. The required model statement is MODEL Y = A . The SAS output consists of an ANOVA table with the p-value of the test. If a p-value of 0.35 was obtained, the conclusion would be: "There is no evidence ($p = 0.35$) that the two fertilizers perform differently."

Example 2: Suppose further that an experiment had treatments with the following levels: a control (no fertilizer), a standard fertilizer, and a new fertilizer applied at two different amounts. The experimenter might have the following questions:

1. Is the control different from the other treatments?
2. Is the standard fertilizer different from the new fertilizer?
3. Are the two different levels for the new fertilizer producing different responses?

These questions could be answered with the following contrasts:

<u>Treatment</u>	<u>Contrasts Coefficients for Question</u>		
	<u>1</u>	<u>2</u>	<u>3</u>
Control	-3	0	0
Standard	1	-2	0
New: level 1	1	1	-1
New: level 2	1	1	1

PROC GLM with the appropriate contrast statements could be used to test the contrasts. The p-values for the contrasts are computed and reported as part of the SAS output. If the p-value for question 1 was 0.015, then the conclusion would be: "There is strong evidence ($p = 0.015$) that the control is different from the other treatments."

Example 3: Suppose it is believed that the growth rate (Y) is linearly dependent on the amount of fertilizer applied (X). To test this claim, fit a regression line $Y = \beta_0 + \beta_1 X$ and test the hypotheses $H_0: \beta_1 = 0$ vs $H_a: \beta_1 \neq 0$. This can be accomplished using PROC REG with the model statement MODEL Y = X. SAS will compute the estimated values of β_0 and β_1 , perform an F-test and report the p-value under the column "PROB > F". If a p-value of 0.023 was obtained, the conclusion could be stated as follows: "There is strong evidence ($p = 0.023$) that the amount of fertilizer applied affects the growth rate."

Example 4: Suppose one is interested in testing if there is a relationship between tree species and success in seed germination. Count data of the number of germinated seeds from each tree species is available for analysis by a contingency table. SAS has 2 procedures to do this: PROC FREQ can be used for two dimensional tables only while PROC CATMOD is suitable for any dimensional table. If PROC FREQ was used, the statement

```
TABLE SPECIES*SEED/CELLCHI2 CHISQ
```

would produce a two-way frequency table of species by seed with the χ^2 statistics due to each cell printed. The total χ^2 statistic and the corresponding p-value would also be reported. If a p-value of 0.084 was obtained, then the conclusion would be: "There is moderate evidence ($p = 0.084$) that success in seed germination is different for the different tree species."

In short, p-values convey information about the strength of evidence against the null hypothesis and allow an individual to draw a conclusion at any specific level α . Nevertheless, α should not be the only criteria for decision making. When the null hypothesis is not rejected, β , the size of the Type II error should also be reported. A false H_0 could have been missed if β is large in the experiment due to a small sample size or large sampling variability. This topic of Power Analysis will be explored in future pamphlets.

References:

Devore, J.L., 1987. *Probability and Statistics for Engineering and the Science*. 2nd ed. Brooks/Cole Publishing Company.

Stafford, S.G., *A Statistics Primer for Foresters*, Journal of Forestry, March 1985, p 148-157.

CONTACT: Vera Sit or Wendy Bergerud
356-0435 387-5676

 NEW PROBLEM

List the null hypotheses for the three questions posed in example 2.

 PROBLEM FROM BI# 29

In order to compute the 90% confidence interval, the value in the `tinvt` statement should be changed from 0.975 to 0.95.

SAS Output for Problem:

Point Estimate and 90% Confidence Interval of X on a Given Y

Model : $Y = \text{ahat} + \text{bhat} * X$

 AHAT BHAT

 64.2468 -1.01299

----- GROUP=1 -----

 No. of

Y-values used	Mean of Given Y	Estimated X	Lower Limit	Upper Limit
1	40	23.9359	37.8890	11.6352
