Appendix E – Certification Test Plan for AFC/MFC Test Stand

1. Introduction

This document describes the certification test plan to qualify the new AFC/MFC test stand (hereinafter referred to as New Stand), to be located in Building 3907 Fuel Accessories Testing shop, and to perform and meet the production testing requirements of all listed Units Under Test (UUTs) per applicable T.O.(s). Note certain items will run a "Full" correlation process while others will run a "Functional" correlation process.

The New Stand will be correlated to prove that performance meets, or exceeds, the existing production test stands and applicable T.O. requirements. The New Stand will be approved for production testing of the End Items listed in Table 1 if the Master Stand data and the New Stand data correlate as follows

2. Full Correlation Process

2.1 Master Stand and New Stand data full correlate procedure is as follows:

2.1.1 The final acceptance test will be run at least three (3) times, per T.O., on the Master Stand to obtain correlation data. Data will be checked for outliers utilizing the Weisberg t-Test. See section 2.2 below for a brief description of testing for outliers. If an outlier is found, results from an additional run will be utilized to ensure that three samples are available for each test point. If an outlier is still indicated, it will be considered normal data scatter and all four runs will be used in the 95% confidence interval analysis.

2.1.2 Repeat the steps in section 2.1.1 above but testing on the New Stand instead.

2.1.3 Correlation between the two data sets will be analyzed utilizing the MOSI Method with confidence intervals of 95%. See section 2.3 below for a brief description of analysis method. See Attachment 4 for detailed description of analysis method.

2.1.4 The results from the full correlation methods will be documented in an Excel spreadsheet. See Attachment 1 for an example Certification Test Worksheet.

Table 1 - UUTs to be Full Correlation Tested

UUT	Part number	Technical Order
F100-220 AFC	441422-7	6J3-2-34-2-2
F100-229 AFC	790100	6J3-2-36-3
F100-220 MFC	441610-21	6J3-4-118-2-2
F100-220 MFC	441396-21	6J3-4-118-2-2
F100-220 MFC	442645-23	6J3-4-118-2-2
F100-229 MFC	441721-10	6J3-4-118-2-2
PS2 Computer Bracket (for 220 MFC)	(Component)	6J3-4-118-2-2

2.2 Weisberg t-Test for Outliers

All correlation data will be tested for Outliers at an Alpha Error Rate of 5%. All data will be checked for outliers at the time of data generation from the test stand and at the time of data correlation. See Attachment 3 for additional information.

2.3 Mosi Method

The Mosi Method for comparing two means with a 95% Confidence interval involves the following steps (See Attachment 4 for additional information):

- Average the three samples from each stand
- Calculate the Standard Deviation of the three samples from each stand
- Calculate the 95% confidence intervals for each mean of stands

The data from each stand is statistically comparable if the intervals overlap such that each interval overlaps with the mean value of the other interval as shown in the diagram below. In our case there is a potential bias that must be taken into account to properly apply the MOSI technique. This bias is the measurement uncertainty of the test stand instrument.

2.4 Example

Test stand A measures an average flow (3 readings) of 960 pph and Test stand B measures an average flow (3 readings) of 1040 pph. The measurement uncertainty for the flow measurement is ± 50 pph. For test stand A we can state the true average flow measurement is between 910 and 1010 pph and for test stand B we can state the true average flow measurement is between 990 and 1090 pph. Due to this bias we must include the measurement uncertainty as a range of acceptable mean values to properly apply the MOSI technique. This is graphically displayed on the following page.

3. Functional Correlation Test

3.1 Master Stand and New Stand data functional correlate procedure is as follows:

3.1.1 The final acceptance test will be run one (1) time, per T.O., on the Master Stand to obtain baseline data. Data will be checked for failing test point outliers. If an outlier is found, results from an additional run will be utilized to ensure UUT is at or close to A-Condition. If an outlier is still indicated, it will be considered best case and final run will be used in the analysis.

3.1.2 Repeat the steps in section 3.1.1 above but testing on the New Stand instead.

3.1.3 The results from the functional correlation method will be documented in an Excel spreadsheet. See Attachment 1 for an example Certification Test Worksheet.

Table 2 - UUTs to be Functional Correlation Tested

UUT	Part number	Technical Order
LRU-AFC Fill Switch Module	351014A	6J3-2-34-8-1
LRU-AFC Core Resolver	(Component)	6J3-2-34-8-1
LRU Sequencing Resolver	351019A	6J3-2-34-8-1
AFC Reset Manifold	2671164	6J3-2-34-8-1
AFC Resolver Assembly	351013	6J3-2-34-8-1
F100 PLA Resolver	351006	6J3-4-118-2-2
F100 MV Resolver	351007	6J3-4-118-2-2
PS2 Computer Bracket (for 220 MFC)	(Component)	6J3-4-118-2-2
PLA Rate Limiter (for 220 MFC)	(Component)	6J3-4-118-2-2
High Pressure Body (for 220 MFC)	(Component)	6J3-4-118-2-2

4. Test Report

After the correlation data runs are completed a preliminary test report will be submitted to Government Technical POC. The System Program Office Cognizant Engineering Authority will also be provided the test report for review. Following Cognizant Engineering Authority approval, contractor will be notified of approval of the New Stand. See Attachment 2 for an example Certification Test Report.

5. Attachments

Attachment 1
Example Certification Testing Worksheet Example.xls

Attachment 2
Example Certification Test Report.pdf

Attachment 3
Weisberg-t Test for Outliers.pdf
T-Distribution Table.pdf
Analysis of two data samples using the MOSI Method.pdf

Attachment 1

Correlation Worksheet

Part Name	Augmenter Fuel Control	Equipment Specialist	Kenneth Ham/736-3482
P/N and S/N	1459M17G04/GAT1G058	Cognizant Engineer	Mike Babb/736-3517
T.O. Number	6J3-2-33-3	Test Mechanic	
C/N		Production Engineer	
Test Specification/Paragraph/Table	Log Sheets: T.O. Section 7.10 (b)	Test Dates	

Metering Valve Minimum Stop (Core Cut-In Overlap)

Test Point Information								Master Stand (3C3973G03-0C001/0303)													Modified Stand (3C3973G03-0C002/0403)												
Test Description	Paragraph	Parameter Name	Parameter Units	Test Spec MIN Limit	Test Spec MAX Limit	Test Spec Nominal Value	Req'd Instrument Accuracy	Run #1 08/12/10	Run #2 08/12/10	Run #3 08/13/10	Run #4 08/13/10	Run #5 08/13/10	Run #6 08/13/10	Mean	Mean Less Acc.	Mean Plus Acc.	Std. Dev.	95% Conf. MIN	95% Conf. MAX	95% Conf. Correlation Master => Modified	Suspect Outlier?	Set-Up Errs or OOL Reads	Run #1 08/20/10	Run #2 07/26/10	Run #3 03/21/11	Run #4 08/09/10	Run #5 03/21/11	Run #6	Mean	Mean Less Acc.	Mean Plus Acc.	Std. Dev.	95% Conf. MIN	95% Conf. MAX	95% Conf. Correlation Modified => Master	Suspect Outlier?	Setup Errs or OOL Reads
Test Condition	7.10(b) 1.000	PR (062JA | MA)	psig	950	1050	1000	±10.00	1005	1003	1003	1001	1001											998	1001	964	998	967									
Test Condition	7.10(b) 1.000	PB1 (062JC | NB)	psig	90	110	100	±0.5	96	101	98	96	97											100	102	100	103	99									
Test Condition	7.10(b) 1.000	CDP (060MA|SA|PA)	psig	365	375	370	±0.8	368	370	369	371	368											370	370	370	370	370									
Test Condition	7.10(b) 1.000	P5 (062PA | VA)	psig	185	205	195	±5.0	196	196	202	202	198											190	188	191	196	197									
Test Condition	7.10(b) 1.000	P6 (062PB | VB)	psig	195	245	220	±5.0	197	244	233	244	215											216	220	219	226	244									
Test Condition	7.10(b) 1.000	P7 (062PC)	psig	165	215	190	±5.0	195	207	170	190	182											194	201	196	184	176									
Test Condition	7.10(b) 1.000	PC (062JB | NA)	psig	275	295	285	±0.5	287	287	295	287	282											288	290	288	282	280									
Test Condition	7.10(b) 1.000	Torque Motor Current	mA	-5.0	0.0	-2.5	±0.1	0.0	0.0	0.0	0.0	0.0											0	0	0	0	0									
Test Result	7.10(b) 1.000	Feedback Volt/Volt Rdg	**v/v**	0.4035	0.4065	**0.4050**	**±0.001**	0.4060	0.4059	0.4060	0.4059	0.4060		**0.4060**	**0.4050**	**0.4070**	**0.0001**	**0.4049**	**0.4071**	*Pass*	*No*	*OK*	0.4064	0.4063	0.4042	0.4061	0.4041		**0.4050**	**0.4040**	**0.4060**	**0.0012**	**0.4025**	**0.4075**	*Pass*	*No*	*OK*

Attachment 2

CERTIFICATION TEST Report

for

Part Types and Numbers

on

Test Stand Type

PN: XXXX

LOCATION: BUILDING 2210

REPORT DATE: 1 APRIL XXXX

(Part Type) Certification Test Report

Introduction

This document briefly describes the certification test reports testing results for qualifying the XXX Test Stand, located in Building XXXX.

Background

The new (Test Stand Type) Test Stand (part number) was delivered under contract # XXXXX.

System Overview

General description of equipment and basis of design.

System Calibration and Accuracies

Test Stand calibration will be performed and certified by Tinker PMEL Technicians IAW T.O. 00-20-14. Calibration accuracies are identical to P/N XXXX.

Part Number For Correlation Testing

The Part Number listed in TABLE 1 was utilized for the correlation study.

TABLE 1 Part Number Used For Correlation Testing

Part Number	Noun	T.O.	Test Type
XXX	XXX	XXX	XXX
XXX	XXX	XXX	XXX

Correlation Process

1. A serviceable (part type(s)) listed in TABLE 1 was run five times on the Master Test Stand (test stand part #) to obtain baseline correlation data. The augmenter fuel controller was then run five times on the New Test Stand (test stand part #) and data recorded.
2. Data results were checked for outliers utilizing the Weisberg t-Test (See Attachment M – Statistical Correlation and Validation Method). If an outlier was flagged by the t-Test an additional run was made and replaced the outlier. If an outlier was still indicated after replacing the suspected outlier data point, it was considered normal data scatter and all six runs were used in developing the 95% confidence interval analysis.
3. The results from the correlation methods were documented in Excel spreadsheets. (See Attachments A thru L – (Part Type) Correlation Worksheets)
4. Correlation between the two data sets was accomplished utilizing the MOSI Method with a confidence interval of 95%. (See Attachment M – Statistical Correlation and Validation Method).

Signature Sheet

The undersigned certify XXXX Test Stand, Test Stand Part Number XXXXXXXX meets or exceeds all measurement and test requirements of Technical Order XXXXXXX.

_____
Contractor Signature

_____
Cognizant Engineer of Part

Attachments:

- Attachments A thru L - (Part Type) Correlation Worksheets
- Attachment M - Statistical Correlation and Validation Method.pdf

Attachment 3

Weisberg t-Test for Outliers

All correlation data will be tested at an Alpha Error Rate of 5% for Outliers. All data will be checked for Outliers at the time of data generation from the test stand and at the time of data correlation.

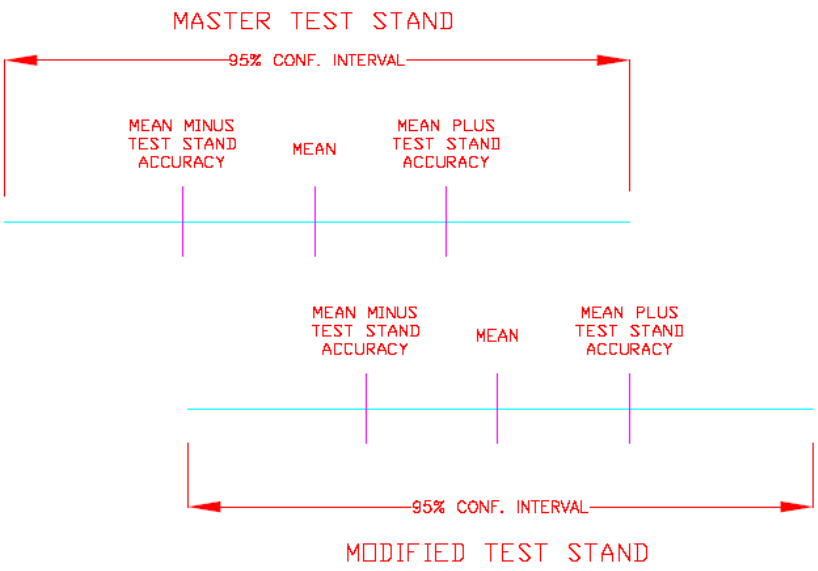
Mosi Method

The Mosi Method (See Attachment 5 for additional information) for comparing two means with a 95% Confidence interval involves the following steps:

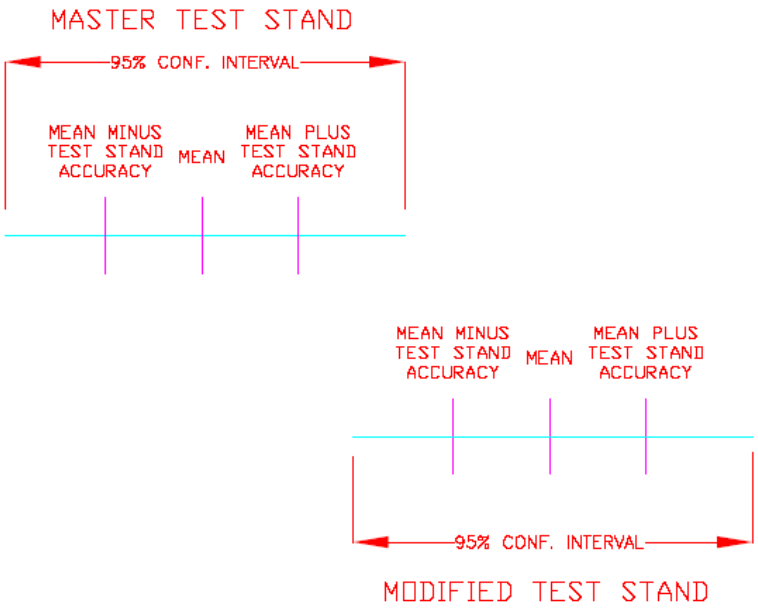
- Average the three samples from each stand
- Calculate the Standard Deviation of the three samples from each stand
- Calculate the 95% confidence intervals for each mean of stands

The data from each stand is statistically comparable if the intervals overlap such that each interval overlaps with the mean value of the other interval as show in the diagram below. In our case there is a potential bias that must be taken into account to properly apply the MOSI technique. This bias is the measurement uncertainty of the test stand instrument. For example:

> Test stand A measures an average flow (3 readings) of 960 pph and Test stand B measures an average flow (3 readings) of 1040 pph. The measurement uncertainty for the flow measurement is ± 50 pph. For test stand A we can state the true average flow measurement is between 910 and 1010 pph and for test stand B we can state the true average flow measurement is between 990 and 1090 pph. Due to this bias we must include the measurement uncertainty as a range of acceptable mean values to properly apply the MOSI technique. This is graphically displayed below.



Example of two stands that are statistically comparable



Example of two stands that are not statistically comparable

Confidence intervals

- ΔX at a given confidence level (say 95%) implies that the true value μ will be found within $\pm \Delta X$ of the calculated mean

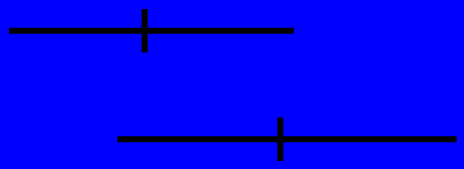
$$\mu = \bar{x} \pm \frac{t_{p,v} s}{\sqrt{N}}$$

$$\left(\bar{x} - \frac{t_{p,v} s}{\sqrt{N}} < \mu < \bar{x} + \frac{t_{p,v} s}{\sqrt{N}} \right)$$

- s = standard deviation **between individual values**
- t = Student t value at a given probability (See Student T Distribution)
- $\bar{x}$ = Mean of data samples
- N = Number of Samples

Comparing two means (unpaired data)

- Mosi method
  - Calculate the confidence interval for each mean
  - Compare the confidence intervals
  - *The results are statistically comparable if the intervals overlap such that **each interval overlaps with the mean value of the other interval** as shown in the diagram below.*



Demonstrating the Consistency of Small Data Sets

Application of the Weisberg *t*-test for Outliers

Robert J. Seely, Louis Munyakazi, John Haury, Heather Simmerman, W. Heath Rushing, and Thomas F. Curry

Determining whether a data point is an "outlier" — a result that doesn't fit, that is too high or too low, that is extreme or discordant — is difficult when using small data sets (such as the data from three, four, or five conformance runs). The authors show that the Weisberg *t*-test is a powerful tool for detecting deviations in small data sets.

A member of BioPharm International*'s editorial advisory board, corresponding author* ***Robert J. Seely*** *is the associate director of corporate validation, corporate QA;* ***Louis Munyakazi*** *is project biostatistician, six sigma;* ***John Haury*** *is associate director, six sigma;* ***Heather Simmerman*** *is associate director, quality analytical laboratories; and* ***W. Heath Rushing*** *is a quality engineer III at Amgen Inc., 4000 Nelson Road, Longmont, CO 80503, 303.401.1586, fax 303.401.2820, rseely@amgen.com.* ***Thomas F. Curry*** *is a technical fellow at Northrop Grumman Information Technology, 5450 Tech Center Drive, Colorado Springs, CO 80919.*

The attempt to define an "outlier" has a long, diverse history. Despite many published definitions, statisticians in all fields are still interested in objectively determining whether a data point is consistent with the rest of the data set or is, after all, an outlier, signifying a deviation from the norm. For biopharmaceutical companies, the need to evaluate whether a data point is an outlier — inconsistent with the rest of a small set of data — is important in validating process consistency.

The power of a statistical tool increases as sample size (n) increases. So, low power outlier tests used on small data sets — such as the production data derived from three to five conformance runs — have relatively high beta (β) or Type 2 errors. These errors mean there is a high chance of leaving deviant results undetected, making these tests inappropriate for pharmaceutical or biopharmaceutical applications.

The z-test is the most powerful outlier test (the most able to detect a discordant datum) if the data are normally distributed and the standard deviation is known or can be accurately estimated. But z-tests usually require large data sets. Conformance runs from early commercial lots usually produce small data sets, and the standard deviation of the population is not usually known. We show, using representative biopharmaceutical process validation data, that the Weisberg t-test is a powerful outlier test at small values of n. It has a low β error rate in detecting deviations from the mean. Therefore, the Weisberg t-test is suitable for objectively demonstrating consistency of production data.

Process Validation Data

During process validation, process consistency is typically demonstrated in three to five conformance runs. When historical data are available, even if those data are from a different production scale, statistical comparisons are relatively straightforward. Those data can be used to set acceptance criteria for validation runs at a new production scale.

Control charts. When about 15 lots have been produced at commercial scale, control charts — which present a picture of a process and its variation over time — are useful for evaluating process stability. Our choice of 15 lots for calculating control limits is a balance between the extreme uncertainty of limits based on few data and the diminishing value of each new data point in further decreasing that uncertainty. An individuals control chart based on 15 lots has 8.9 effective degrees of freedom (df), which is sufficient to reduce the uncertainty in the limits to about $\pm$23%. Achieving $\pm$10% uncertainty requires about 45 degrees of freedom, which requires more than 70 individual values (1).

Outlier tests and errors. Until there are 15 lots, the most useful method to statistically evaluate a data point that seems to be anomalous is the Weisberg t-test (2,3). The Weisberg t-test can be used for data sets larger than 15 values as well. The other tests we evaluated for application in small data sets were the Dixon (4) and the Grubbs (5). The latter is also known as the Extreme Studentized Deviate (ESD) (6). The Weisberg t-test can distinguish between normal process variation and a process aberration that yields an outlier. For example, at an alpha (α) error (calling something an outlier when it isn't one) of 0.05, the β error is only 0.31, and thus the Weisberg t-test is the most powerful test available for small data sets among those considered (Figure 1).

Discordant Observations

Until recently, the *U.S. Pharmacopeia* (USP) did not address the treatment of chemical test data containing discordant observations. Indeed, this "silence" was interpreted to mean a "prohibition" during

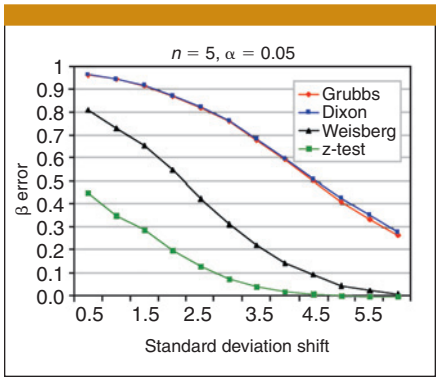


Figure 1. The operating characteristic curves for a variety of outlier tests are created by repeatedly drawing four values from a population of known mean and S, with a fifth value taken from a population with a known shift in mean.

the *United States v. Barr Laboratories, Inc.* case (7). Judge Wolin's ruling in that case indicated the need for such guidance (8), and in 1999, a new monograph was previewed in *Pharmacopeial Forum* (9). That monograph states that when appropriately used, outlier tests are valuable tools for analyzing discordant observations.

The discussions in the *Barr* case and in the USP monograph suggest the appropriateness of using an outlier test to disregard a data point. In this article, we use such a test — the Weisberg *t*-test — to objectively identify an outlier as part of a statistical evaluation of small data sets. For process validation purposes, if the Weisberg *t*-test identifies no outliers, the data can be claimed to be consistent based on an objective statistical method.

This article describes the application of the Weisberg *t*-test to data from five conformance runs. We examine the ability of the test to demonstrate process consistency. Subsequent uses of this test would include checking a *suspect* data point from lot six with the previous five, or lot seven from the previous six, for instance. The Weisberg *t*-test could also be used during a retrospective review of data sets. For example, an earlier value may stand out as a possible outlier, but only after subsequent data show a pattern that distinguishes it as a possible outlier. As standard practice, we advocate an investigation of the causes of such statistical differences.

As mentioned, at 15 data points, the individuals control chart for each point becomes the preferred tool for detecting discordant observations and for showing process consistency (defined as the absence of discordant observations). If the data are available in subgroups, then an averages control chart is preferred.

Testing for a Single Outlier

In this article, we refer to an outlier as a datum that appears not to belong to the same group as the rest of the data. That datum measurement may seem either too large or too small in relation to the general pattern of the rest of the data. The method we propose applies to a single outlier, and is similar to the traditional *t*-calculated (t_{calc}) form of the general *t*-test statistic (Equation 1) (10,11).

$$t_{calc} = \frac{\left\{ Estimate\ \ of\ the\ difference \right\}}{\left\{ Standard\ error\ \ of\ the\ difference \right\}} \quad [1]$$

The test hypothesis (H_o or null) can be stated as: The suspected value is not an outlier. Its alternative (H_a or alternative) is stated as: The suspected value is an outlier.

Working with reduced data. The entire set of data should not be used to estimate the standard error (SE). Such estimates would be biased if the suspected outlier were included. The estimate of variation would be inflated, and the estimate of the arithmetic mean would be biased toward the outlier.

The logic of the Weisberg *t*-test. After computing the estimates without the suspected outlier, the Weisberg *t*-test statistic for the suspected outlier (denoted by y_i) is given in Equation 2,

$$t_{calc} = \frac{\left(\frac{n-1}{n}\right)^{\frac{1}{2}} \times \left(y_i - \bar{y}_{-i}\right)}{s_{-i}} \quad [2]$$

where n is the sample size, $\bar{y}_{-i}$ denotes the computed sample mean, s_{-i} is its standard deviation after the withdrawal of y_i, the suspect outlier.

The logic of the Weisberg *t*-test is that the numerator $(y_i - \bar{y}_{-i})$ compares the mean value $\bar{y}_{-i}$ to the suspected outlier value y_i. Furthermore, the denominator s_{-i} is the classic sample standard deviation.

In Equation 2, the factor denoted by

$$\left(\frac{n-1}{n}\right)^{\frac{1}{2}}$$

adjusts the calculated *t*-value (t_{calc}) downward and is more conservative for small samples. Specifically, the above factor is identical to

$$\left(1 + h\right)^{\frac{1}{2}}$$

where h is the leverage matrix (3); that is

$$h = \frac{1}{n-1}$$

because the data are reduced by one observation. The estimated SE, of the mean $\bar{y}_{-i}$ is

$$SE\left(\bar{y}_{-i}\right) = s_{-i}\left(1 + h\right)^{\frac{1}{2}} \quad [3]$$

which makes Equation 2 equivalent to

$$t_{calc} = \frac{\left(y_i - \bar{y}_{-i}\right)}{SE\left(\bar{y}_{-i}\right)} \quad [4]$$

The t_{calc} (as found above) is then compared to percentiles of a *t*-distribution at the α significant level with $(n-2)$ degrees of freedom.

Probability and degrees of freedom. Table 1 shows the *t*-critical (t_{crit}) values at three different α values, in which the *df* are two less than the sample size. If the absolute value of t_{calc} is less than t_{crit}, the point is not an outlier. Table 1 can be generated (in Microsoft Excel or another spreadsheet program) for other values of α (alpha error rate) and degrees of freedom using the inverse of the Student's *t*-distribution (TINV) function. The TINV function requires two arguments. They are: "the probability associated with a one-tailed *t*-distribution" and the "degrees of freedom." The critical *t*-values in the body of Table 1 are derived using the Excel function: TINV(reference to value at the top of the column as a proportion, times two; and the value for the degrees of freedom from the first column where degrees of freedom is two less than the number of samples). In Excel, the TINV function gives the *t*-value for two tails, placing one-half of the α value in each tail, whereas the Weisberg *t*-test is a one-tail test. That is why the α values must be multiplied by two when using the TINV spreadsheet function. When using a published two-sided Student *t*-table, the results are obtained by shifting one column to the right, that is, by using $(n-2)$ degrees of freedom, as shown in Table 1.

Identifying a Biotech Outlier

As a real-world example, we use a data set from monitoring a large chromatography column in a recombinant protein purification process. Table 2 presents a representative set of such data. Step yield (percent recovery), target protein concentration, a purity assay,

Alternative Methods for Determining t_{calc}

ANOVA Table. A one-way analysis of variance model (Alternate 2) delivers the same results as the Weisberg *t* or regression-based tests: the $F_{calc} = t^2_{calc}$; also the *p*-values are identical (Table 3).

Source	Degrees of Freedom	Sum of Squares	Mean Square	F_{calc}
Model	1	SS_{model}	MS_{model}	$\frac{MS_{model}}{MSE}$ [a]
Error	$n-2$	SS_{error}	MS_{error}	
Corrected total	$n-1$	SS_{total}		

[a]MSE is mean square error.

SAS Codes for the Alternative Methods

The Data

```
data o;
input y @@;
x=0;
if _n_=3 then x=1;
datalines;
17.5   17.4   30.2   22.2   27.0
;
run;
proc print;
run;
```

Alternative 1: Linear Regression

```
proc reg alpha=.1;
A: model y=x;
*test H0: b=0;
B: model y=t/influence;
*look for R-Studentized Residual;
output out=hat h=hmatrix;

title3 Method Uses ALL the Data;
title4 Simple Linear Regression Model;
run;
```

Alternative 2: One-Way ANOVA

```
proc glm data=o alpha=.1;
class x;
model y = x/ss3 solution;
output out=g h=hamtrix;

estimate 'Estimate of mu' intercept 1 x 1;
estimate 'Estimate of outlier' intercept 1 x 0 1;
estimate 'Weisberg test' x -1 1;
title3 Method Uses ALL the Data;
title4 Through Linear Model and Contrast;
run;
```

Two alternative approaches to test for outliers include the regression approach (Alternative 1) and the ANOVA approach (Alternative 2). The results of the two alternatives are compared with the Weisberg *t*-test in Table 3. The entire data set ($n=5$) is used for these methods, but the results are identical to those obtained using Equation 2 or Equation 4.

Alternative 1: The Regression Approach

An outlier test similar to Equation 2 exploits the full data set by testing the hypothesis that $\beta_1=0$ using a simple linear model:

$$y = \beta_0 + \beta_1 X^i \text{ with } var(y) = \sigma^2 \quad [5]$$

in which *y* is the expected value given X^i. The term X^i is coded 1, if *y* is the suspected outlier ($y=y_i$) and 0 otherwise. In this model, β_0 estimates the overall mean, and β_1 represents the deviation from the mean for the rest of the data. The *t* statistic for testing $\beta_1=0$ against a two-sided alternative is the appropriate statistic to use (12). Under the assumption of normal error, the *t* is a Student *t* with $n-k-1$ degrees of freedom, in which $k=1$ (due to β_1). Therefore,

$$t_{calc} = \frac{b_1}{SE(b_1)} \quad [6]$$

in which b_1 and $SE(b_1)$ are sample estimates of β_1 and its standard error (SE).

Alternative 2: The One-Way ANOVA Model

A similar coding of the full data leads to the same test through the use of one-way analysis of variance (ANOVA). The model is:

$$y = \mu + \alpha \quad with \ var(y) = \sigma^2 \quad [7]$$

in which *y* is the expected value, μ is the overall mean, α represents two classes or categories defined by 0 and 1 depending on whether the observation is a suspected outlier (=1) or not (=0). The degrees of freedom are $(n_\alpha-1)$, in which n_α are the two levels of α, therefore $(n_\alpha-1)=(2-1)$. Consequently, the degrees of freedom for the error is $(n-1) - (n_\alpha-1)$. The ANOVA table is provided below.

The *df* column in the ANOVA table defines the degrees of freedom; the F_{calc} is equivalent to t_{calc} and have identical probability of discerning an outlier (the *p*-value); that is, F_{calc} equals the t^2_{calc} obtained in the Weisberg *t*-test and in the regression approach. Moreover, estimates of $\bar{y}_{-i}$ and *y* are obtained by applying estimable functions, that is

$$\bar{y}_{-i} = m + a_0 \text{ and } y_i = m + a_1 \quad [8]$$

where *m*, a_0, and a_1 can be obtained from the solution vector of the model in Equation 7. The elements of the solution vector — *m*, a_0, and a_1 — represent nonunique estimates of the intercept (*m*), the effect of observations without the suspected value (a_0), and the effect of the suspected value (a_1). A test identical to the Weisberg *t*-test and the regression-based test is obtained by computing the difference between the two estimable functions in Equation 8. The resulting difference is also estimable (12). Thus,

$$t_{calc} = \frac{(a_0 - a_1)}{SE(a_0 - a_1)} \quad [9]$$

provides an equivalent test to the Weisberg *t*-test and the regression-based test (3,5). The standard error $SE(a_0-a_1)$ is

$$SE(a_0 - a_1) = \sigma\sqrt{\left(\frac{1}{n} + 1\right)} \quad [10]$$

The Weisberg *t*-test, the regression-based test, and the ANOVA model are similar because in all three methods, the same quantity (in absolute terms) is represented by the estimated slope β_1 of Equation 5, the numerator (a_0-a_1) in Equation 9, and the numerator

$$\left(y_i - \bar{y}_{-i}\right)$$

of Equation 4. All three methods also have the same SE. The results of the outlier tests of the three methods are compared in Table 3 using the data from Table 2.

Because of their simplicity, the above calculations can be performed in a spreadsheet package that has even limited statistical capability. The SAS code needed to run these methods is listed in the box to the left (13). They can also be obtained from Louis Munyakazi, louism@amgen.com.

Table 1. The t_{crit} values for three different levels of α errors and the degrees of freedom (two less than the sample size)

Degrees of Freedom[a]	Alpha Error Rate 0.01	0.05	0.10
3	4.541	2.353	1.638
4	3.747	2.132	1.533
5	3.365	2.015	1.476
6	3.143	1.943	1.440
7	2.998	1.895	1.415
8	2.896	1.860	1.397
9	2.821	1.833	1.383
10	2.764	1.812	1.372
11	2.718	1.796	1.363
12	2.681	1.782	1.356
13	2.650	1.771	1.350
14	2.624	1.761	1.345
15	2.602	1.753	1.341
16	2.583	1.746	1.337
17	2.567	1.740	1.333
18	2.552	1.734	1.330

[a]Degrees of freedom are *n*–2 of sample size.

Table 2. A representative set of data from the first five lots of a purification process for a recombinant protein in a large chromatography column; the parameters are the primary indicators of step consistency.

Parameter	Lot 1	Lot 2	Lot 3	Lot 4	Lot 5
Step time (h)	50	49	48	51	51
Step yield (%)	82	83	89	88	89
Concentration (g/L)	13.8	13.6	14.1	13.8	14.0
Purity (%)	97.1	97.2	97.6	97.3	97.5
HCP (ppm)	17.5	17.4	30.2	22.2	27.0

Table 3. A comparison of the Weisberg *t*-test with two other methods for obtaining identical t_{calc} values: the regression-based method (Alternate 1) and the ANOVA-based method (Alternate 2); the host cell protein (HCP) observations for testing step consistency are the responses being tested.

Method	Number of Observations	Estimate ± Standard Error	Calculated *t*	*p*-value
Weisberg *t*-test (reduced data)	$(n-1) = 4$	21.03 ± 4.57	1.796	0.1704
Alternate 1 Regression test (full data)	$n = 5$	9.18 ± 5.11[a]	1.796	0.1704
Alternate 2 ANOVA test (full data)	$n = 5$	21.03 ± 2.28[b] 30.20 ± 4.57[b] 9.18 ± 5.11[b]	1.796	0.1704

[a]Represents estimates of β_1 (deviation from the mean of the $n-1$ data)
[b]Represents estimates of $m+a_0$, $m+a_1$, and a_0-a_1 (mean of reduced data, suspected outlier, and their difference, ± corresponding SEs)

host cell protein (HCP) concentration, and processing time are the primary indicators of step consistency. The data appear to be consistent across the five lots, except in Lot 3, the HCP is apparently high and might be inconsistent with the other four data points.

The Weisberg t_{calc} for HCP in our example is 1.796. Comparing that number with the t_{crit} values (Table 1), for $n=5$, $\alpha=0.05$, the t_{calc} is less than the t_{crit} (2.353), therefore the data point is not an outlier, and the five data points are consistent. So the subjective judgment used to decide that the data point might be discordant is followed by the application of a statistical tool to give an objective assessment.

Choosing an α value of 0.05 means that when the process actually has no outliers, we are willing to accept a 5% chance of a false positive — a 5% chance that a point identified as discordant by the Weisberg *t*-test is not, actually, an outlier. Accepting that rate means accepting unnecessary investigations 5% of the time. If the α value is reduced to avoid those investigations, the β value rises, which means false negatives — the test fails to identify a discordant value. In our application, a β error occurs when the test fails to detect an outlier when one is actually present. We choose to set α at 0.05 and are willing to perform more frequent investigations (as a result of false positives) to keep the β error rate low. At $\alpha=0.05$, $n=5$, the β error is a reasonable 0.31 for detecting a shift of three standard deviations (Figure 1). That rate is significantly less than the more commonly used outlier tests (Dixon and Grubbs), which yield β errors of approximately 0.77 as can be seen in Figure 1 (4,5).

The set of operating characteristic (OC) curves in Figure 1 shows a variety of outlier tests constructed by simulating a data set of 5,000 from which four samples were drawn. A fifth datum was randomly taken from a data set, which was shifted by a given number of standard deviations. To detect a standard deviation change of three, the *z*-test (the basis for control charts) is clearly the most sensitive, with a β error of 0.08. For the purposes of comparing outlier tests, the *z*-test is presented here as a one-sided test; the two-sided *z*-test is the basis for control charts. The control chart, however, requires a large data set or a good estimate of the variance of the data. When those conditions are available, an individuals control chart (or a control chart of averages) is the recommended method. When a large data set or a good estimate of the variance is not available — for five conformance runs with little relevant data from previous scales, for example — the Weisberg *t*-test is clearly the next best available tool.

The Bonferroni correction is used for some statistical comparisons. For example, it is used for multiple comparisons (the "family" of comparisons) by dividing the Type 1 error among all comparisons, so that the overall Type 1 error rate of the family does not exceed a desired level. In our example, we use a single hypothesis test for one visually suspected outlier, rather than testing a hypothesis of no outliers by performing multiple tests on every data point versus the remaining set. Because multiple comparisons are not contemplated in our example, we don't use the Bonferroni correction.

One-sided or two. A final point needs to be made about the one-sided versus the two-sided Weisberg *t*-tests for outliers. Because an outlier is initially detected as being the farthest from the central tendency (the mean) of the data, the outlier will be either

Continued on page 58

AD INDEX

Company	Page	Info #	Phone	Fax	Web Site
ABBOTT BIORESEARCH CENTER	19	22	508.849.2785	508.755.8506	www.abbott.com/abbottbioresearch
ADVANCED RESEARCH TECHNOLOGIES (ART)	7	3	514.832.0777	514.832.0778	www.art.ca
AGILENT TECHNOLOGIES INC.	17	9	800.227.9770	800.519.6047	www.agilent.com
AMERSHAM BIOSCIENCES	39	39	732.457.8000	732.457.0557	www.amershambiosciences.com
ASME INTERNATIONAL	54	43	800.843.2763	212.591.7143	www.asme.org/education
AVECIA	18	17	+44.161.721.1831	+44.161.721.5319	www.avecia.com/biotech
BIACORE	43	25	+49.18.675.700	+49.18.150.110	www.biacore.com
BIOENGINEERING AG	6	5	+41.55.256.8111	+41.55.25.68256	www.bioengineering.ch
BIO-RAD LABORATORIES	68	20	800.424.6723	510.741.5800	www.bio-rad.com/process
BIORELIANCE	37	31	301.738.1000	301.738.4036	www.bioreliance.com
BROADLEY JAMES, INC.	57	7	949.829.5555	949.829.5560	www.broadleyjames.com
CHARLES RIVER LABORATORIES	23	32	978.658.6000	978.658.7132	www.criver.com
CUNO INC.	45	23	203.238.8930	203.238.8977	www.cuno.com
DIOSYNTH RTP, INC.	47	30	919.678.4400	919.678.4499	www.diosynth.com
GIBCO INVITROGEN CORPORATION	11	13	800.955.6288	760.603.7229	www.invitrogen.com/gibco
HAMILTON SUNDSTROM	48	45	909.593.3581	909.392.3207	www.hamiltonsundstrand.com
HYCLONE	13	8	800.492.5663	800.533.9450	www.hyclone.com
IBM LIFE SCIENCES	15	4	914.499.1900		www.ibm.com/solutions/lifesciences
IRVINE SCIENTIFIC	25	16	800.437.5706	949.261.6522	www.irvinesci.com
LAUREATE PHARMA LP	2	1	609.919.3400	609.520.3963	www.laureatepharma.com
LONZA BIOLOGICS, INC.	27	15	603.334.6100	603.334.3300	www.lonzabiologics.com
MALLINCKRODT BAKER	21	12	908.859.2151	908.859.9385	www.mallbaker.com
MILLIPORE CORPORATION	5	2	781.533.2117	781.533.3117	www.millipore.com/bioprocess
NEKTAR	49	28	650.631.3100	650.631.3150	www.nektar.com
NIRO SOAVI	51	24	715.386.9371	715.386.9376	www.niroinc.com
NOVA BIOMEDICAL	35	27	800.458.5813	781.894.5915	www.novabiomedical.com
PALL CORPORATION	33	21	800.717.7255	516.625.3610	www.pall.com/biopharmaceuticals
SEROLOGICALS CORPORATION	41	33	678.728.2000	678.728.2020	www.serologicals.com
SPARTA SYSTEMS INC.	9	6	732.203.0400	732.203.0375	www.sparta-systems.com
SWAGELOK COMPANY	31	29	866.737.6246	866.737.6247	www.swagelok.com
TEFEN OPERATIONS MANAGEMENT CONSULTING	53	14	866.8TEFEN8 ext. 135	212.317.0604	www.tefen.com/biobenchmark
UCB BIOPRODUCTS, INC.	67	19	770.437.5500	770.437.5640	www.ucb-bioproducts.com
YSI LIFE SCIENCES	54	26	937.767.7241		www.ysi.com/lifesciences

***Weisberg t-test** continued from page 42*

higher or lower than the mean. The Weisberg *t*-test determines whether the "outlier" is larger if it is to the right of the mean (on a number line) or smaller if it is to the left of the mean (on a number line). The test does not show the differences without reference to the direction of that difference; therefore, the Weisberg *t*-test is a one-sided test, and the resulting t_{crit} values in the table must reflect that.

Alternative methods. Two other methods can be used to obtain identical t_{calc} values. One uses regression (Alternative 1 in Table 3), and one uses analysis of variance (ANOVA) (Alternative 2 in Table 3). These methods are discussed in the "Alternative Methods for Determining t_{calc}" sidebar, and the results from those tests are compared with the Weisberg t_{calc} values in Table 3.

A Superior Tool

The Weisberg *t*-test has a low β error rate (especially when used with a higher α error rate) for small data sets. It is a superior, objective tool for showing consistency within small data sets. As shown in our example, the test fits the needs for evaluating biotechnology process data.

The Weisberg *t*-test can be applied for determining the internal consistency of small data sets and can also be useful in process validation. When validating a process, a protocol with preapproved acceptance criteria is required. For key performance parameters, numerical limits for specific attributes must be defined and met. Typically, however, many secondary parameters may not have predefined numerical limits, but they are still expected to be internally consistent during the validation runs. For example, during scale-up, the mean of a given output parameter can shift up or down, but if that does not affect product quality, the variation may be perfectly acceptable. To validate that a process is performing consistently, the values of that parameter should be similar for three to five runs. The Weisberg *t*-test is a useful tool that adds statistical objectivity to the claim that a process is "consistent." ***BPI***

References

(1) Wheeler, D.J., *Advanced Topics in Statistical Process Control: The Power of Shewhart's Charts* (SPC Press, Knoxville, TN, 1995), p. 185.

(2) Brandt, R., "Comparing Classical and Resistant Outlier Rules," *J. Am. Stat. Assoc.* 85, 1083–1090 (1990). *Note: The error in the formula printed in this reference was corrected in Curry, T.F., "Corrections," J. Am. Stat.* Assoc. *96(456), 1534 (2001).*

(3) Weisberg, S., *Probability and Mathematical Statistics: Applied Linear Regression* 2nd ed. (John Wiley & Sons, New York, 1985).

(4) Dixon, W.J., "Processing Data for Outliers," *Biometrics* 9, 74–89 (1953).

(5) Grubbs, F.E., "Procedures for Detecting Outlying Observations in Samples," *Technometrics* 11, 1–21 (1968).

(6) Bancroft, T.A., "Analysis and Inference for Incompletely Specified Models Involving the Use of Preliminary Test(s) of Significance," *Biometrics*, 20, 427–442 (September 1964).

(7) *United States v. Barr Laboratories, Inc.*, 812 F. Supp. 458 (DNJ 1993).

(8) Kuwahara, S.S., "Outlier Testing: Its History and Applications," *BioPharm* 10(4) 64–67 (April 1997).

(9) "General Information: <1010> Analytical Data — Interpretation and Treatment," *Pharmacopeial Forum* 25(5), 8900–8909 (September–October 1999).

(10) Draper, N.R. and H. Smith, *Applied Regression Analysis* (John Wiley & Sons, New York, 1998), p. 4.

(11) Cook, R.D and Weisberg, S., *Applied Regression Including Computing and Graphics* (John Wiley & Sons, New York, 1999).

(12) Searle, S.R., *Linear Models* (John Wiley & Sons, New York, 1971).

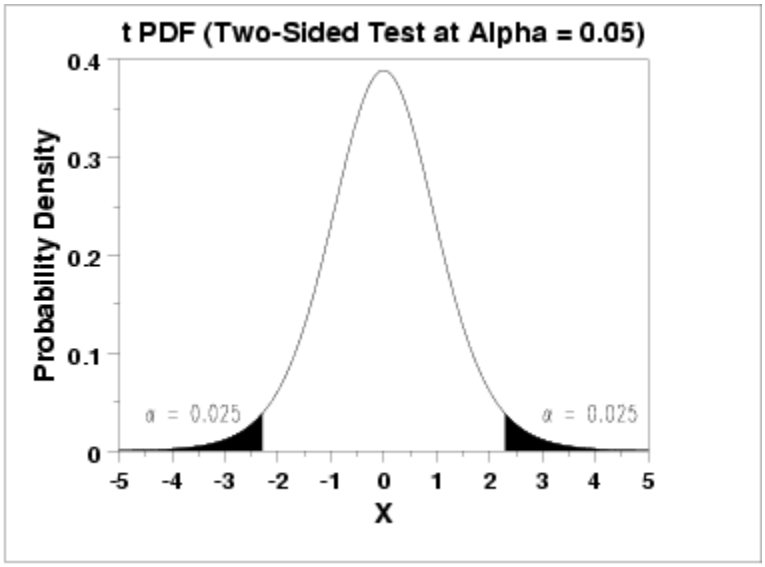
(13) SAS Institute Inc., *SAS/STAT User's Guide*, Version 8.01 (Cary, NC, 1999).

1.3.6.7.2. Upper Critical Values of the Student's-t Distribution

How to Use This Table

This table contains the upper critical values of the Student's *t*-distribution. The upper critical values are computed using the percent point function. Due to the symmetry of the *t*-distribution, this table can be used for both 1-sided (lower and upper) and 2-sided tests using the appropriate value of α.

The significance level, α, is demonstrated with the graph below which plots a t distribution with 10 degrees of freedom. The most commonly used significance level is α= 0.05. For a two-sided test, we compute the percent point function at $\alpha/2$ (0.025). If the absolute value of the test statistic is greater than the upper critical value (0.025), then we reject the null hypothesis. Due to the symmetry of the *t*-distribution, we only tabulate the upper critical values in the table below.



Given a specified value for α:

1. For a two-sided test, find the column corresponding to $\alpha/2$ and reject the null hypothesis if the absolute value of the test statistic is greater than the value of $t_{\alpha/2}$ in the table below.
2. For an upper one-sided test, find the column corresponding to α and reject the null hypothesis if the test statistic is greater than the tabled value.

3. For an lower one-sided test, find the column corresponding to α and reject the null hypothesis if the test statistic is less than the negative of the tabled value.

Upper critical values of Student's t distribution with ν degrees of freedom

	Probability of exceeding the critical value					
ν	0.10	0.05	0.025	0.01	0.005	0.001
1.	3.078	6.314	12.706	31.821	63.657	318.313
2.	1.886	2.920	4.303	6.965	9.925	22.327
3.	1.638	2.353	3.182	4.541	5.841	10.215
4.	1.533	2.132	2.776	3.747	4.604	7.173
5.	1.476	2.015	2.571	3.365	4.032	5.893
6.	1.440	1.943	2.447	3.143	3.707	5.208
7.	1.415	1.895	2.365	2.998	3.499	4.782
8.	1.397	1.860	2.306	2.896	3.355	4.499
9.	1.383	1.833	2.262	2.821	3.250	4.296
10.	1.372	1.812	2.228	2.764	3.169	4.143
11.	1.363	1.796	2.201	2.718	3.106	4.024
12.	1.356	1.782	2.179	2.681	3.055	3.929
13.	1.350	1.771	2.160	2.650	3.012	3.852
14.	1.345	1.761	2.145	2.624	2.977	3.787
15.	1.341	1.753	2.131	2.602	2.947	3.733
16.	1.337	1.746	2.120	2.583	2.921	3.686
17.	1.333	1.740	2.110	2.567	2.898	3.646
18.	1.330	1.734	2.101	2.552	2.878	3.610
19.	1.328	1.729	2.093	2.539	2.861	3.579
20.	1.325	1.725	2.086	2.528	2.845	3.552
21.	1.323	1.721	2.080	2.518	2.831	3.527
22.	1.321	1.717	2.074	2.508	2.819	3.505
23.	1.319	1.714	2.069	2.500	2.807	3.485
24.	1.318	1.711	2.064	2.492	2.797	3.467
25.	1.316	1.708	2.060	2.485	2.787	3.450
26.	1.315	1.706	2.056	2.479	2.779	3.435
27.	1.314	1.703	2.052	2.473	2.771	3.421
28.	1.313	1.701	2.048	2.467	2.763	3.408
29.	1.311	1.699	2.045	2.462	2.756	3.396
30.	1.310	1.697	2.042	2.457	2.750	3.385
31.	1.309	1.696	2.040	2.453	2.744	3.375
32.	1.309	1.694	2.037	2.449	2.738	3.365
33.	1.308	1.692	2.035	2.445	2.733	3.356
34.	1.307	1.691	2.032	2.441	2.728	3.348
35.	1.306	1.690	2.030	2.438	2.724	3.340
36.	1.306	1.688	2.028	2.434	2.719	3.333
37.	1.305	1.687	2.026	2.431	2.715	3.326
38.	1.304	1.686	2.024	2.429	2.712	3.319
39.	1.304	1.685	2.023	2.426	2.708	3.313
40.	1.303	1.684	2.021	2.423	2.704	3.307

Upper critical values of Student's t distribution with ν degrees of freedom

	Probability of exceeding the critical value					
ν	0.10	0.05	0.025	0.01	0.005	0.001
41.	1.303	1.683	2.020	2.421	2.701	3.301
42.	1.302	1.682	2.018	2.418	2.698	3.296
43.	1.302	1.681	2.017	2.416	2.695	3.291
44.	1.301	1.680	2.015	2.414	2.692	3.286
45.	1.301	1.679	2.014	2.412	2.690	3.281
46.	1.300	1.679	2.013	2.410	2.687	3.277
47.	1.300	1.678	2.012	2.408	2.685	3.273
48.	1.299	1.677	2.011	2.407	2.682	3.269
49.	1.299	1.677	2.010	2.405	2.680	3.265
50.	1.299	1.676	2.009	2.403	2.678	3.261
51.	1.298	1.675	2.008	2.402	2.676	3.258
52.	1.298	1.675	2.007	2.400	2.674	3.255
53.	1.298	1.674	2.006	2.399	2.672	3.251
54.	1.297	1.674	2.005	2.397	2.670	3.248
55.	1.297	1.673	2.004	2.396	2.668	3.245
56.	1.297	1.673	2.003	2.395	2.667	3.242
57.	1.297	1.672	2.002	2.394	2.665	3.239
58.	1.296	1.672	2.002	2.392	2.663	3.237
59.	1.296	1.671	2.001	2.391	2.662	3.234
60.	1.296	1.671	2.000	2.390	2.660	3.232
61.	1.296	1.670	2.000	2.389	2.659	3.229
62.	1.295	1.670	1.999	2.388	2.657	3.227
63.	1.295	1.669	1.998	2.387	2.656	3.225
64.	1.295	1.669	1.998	2.386	2.655	3.223
65.	1.295	1.669	1.997	2.385	2.654	3.220
66.	1.295	1.668	1.997	2.384	2.652	3.218
67.	1.294	1.668	1.996	2.383	2.651	3.216
68.	1.294	1.668	1.995	2.382	2.650	3.214
69.	1.294	1.667	1.995	2.382	2.649	3.213
70.	1.294	1.667	1.994	2.381	2.648	3.211
71.	1.294	1.667	1.994	2.380	2.647	3.209
72.	1.293	1.666	1.993	2.379	2.646	3.207
73.	1.293	1.666	1.993	2.379	2.645	3.206
74.	1.293	1.666	1.993	2.378	2.644	3.204
75.	1.293	1.665	1.992	2.377	2.643	3.202
76.	1.293	1.665	1.992	2.376	2.642	3.201
77.	1.293	1.665	1.991	2.376	2.641	3.199
78.	1.292	1.665	1.991	2.375	2.640	3.198
79.	1.292	1.664	1.990	2.374	2.640	3.197
80.	1.292	1.664	1.990	2.374	2.639	3.195
81.	1.292	1.664	1.990	2.373	2.638	3.194
82.	1.292	1.664	1.989	2.373	2.637	3.193
83.	1.292	1.663	1.989	2.372	2.636	3.191
84.	1.292	1.663	1.989	2.372	2.636	3.190
85.	1.292	1.663	1.988	2.371	2.635	3.189
86.	1.291	1.663	1.988	2.370	2.634	3.188
87.	1.291	1.663	1.988	2.370	2.634	3.187
88.	1.291	1.662	1.987	2.369	2.633	3.185

Upper critical values of Student's t distribution with ν degrees of freedom

	Probability of exceeding the critical value					
ν	0.10	0.05	0.025	0.01	0.005	0.001
89.	1.291	1.662	1.987	2.369	2.632	3.184
90.	1.291	1.662	1.987	2.368	2.632	3.183
91.	1.291	1.662	1.986	2.368	2.631	3.182
92.	1.291	1.662	1.986	2.368	2.630	3.181
93.	1.291	1.661	1.986	2.367	2.630	3.180
94.	1.291	1.661	1.986	2.367	2.629	3.179
95.	1.291	1.661	1.985	2.366	2.629	3.178
96.	1.290	1.661	1.985	2.366	2.628	3.177
97.	1.290	1.661	1.985	2.365	2.627	3.176
98.	1.290	1.661	1.984	2.365	2.627	3.175
99.	1.290	1.660	1.984	2.365	2.626	3.175
100.	1.290	1.660	1.984	2.364	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.090