

Statistical Analysis of USDF 2007 Competition Data

Prepared by USDF / USEF Members:

Rebecca Yount, Psy.D.

Ana Diaz

Mary Stydnicki Johnston

27 February 2008

Record of revisions:

Rev. 0 – AED Jan. 23, 2007

Rev. 1 – MSJ 27 Jan 08

Rev. 2 – MSJ 1 Feb 08

Rev. 3 –4 – MSJ 13 Feb 08

Rev. 5 – MSJ 16 Feb 08

Rev. 6 / 6a – MSJ 18 Feb 08

Rev. 7 – MSJ 20 Feb 08

Rev. 8 – AED Feb. 21, 2008

Rev. 9 – AED Feb. 27, 2008-added TOC & expanded summary

Rev. 10 and Final to USEF DC– RY/MSJ, Feb 27, 2008

Rev. 11 / Update – MSJ, Mar 9, 2008

- [Executive Summary](#) 3
- [Introduction](#) 5
- [Summary of Findings, Conclusions, Concerns, and Recommendations](#) 6
 - [FINDING # 1:](#) 6
 - [FINDING # 2:](#) 8
 - [FINDING # 3:](#) 9
 - [FINDING # 4:](#) 10
- [Scope of Data Used](#) 11
- [Types of Analysis and Visualization Techniques Used](#) 11
- [Discussion of Analysis Results](#) 12
 - [Significant statistical variables:](#) 12
 - [Score analysis:](#) 12
- [Specific Observations](#) 13
- [Data](#) 18
 - [Number of scores by Region and Level](#) 18
 - [Analysis of Variance by Judge](#) 20
 - [Regression Analysis: Score versus Region, Level, Judge & Number of Judges](#) 22
 - [One-way ANOVA: Score versus Region](#) 23
 - [One-way ANOVA: Score versus LevelName](#) 24
 - [One-way ANOVA: Score versus CountJudges](#) 25
 - [Histograms for scores by Region & Level](#) 25
 - [Graphical Summary of all scores](#) 34
 - [Cumulative Distribution Function \(CDF\): Rides by score for 1st4, 2nd4, 3rd3, 4th3 & PSG](#) 38

Executive Summary

This analysis was undertaken in response to USEF Proposed Rule, DR119.3 (tracking number 275-07). The stated intent of this rule, “Is to introduce timeline for dressage performance standards system.”

Since no details were provided with DR119.3, the proposed rule created much concern among dressage riders. Additionally, there was confusing information disseminated via various communication means, which created much confusion as to the criteria and intent of the rule. On 1/13/2008, the USEF Board of Directors voted to table this rule until its Annual Meeting in 2009. The USEF Dressage Committee is working on a revised version of the Performance Standards rule, which is due to be released for review and comment some time in 2008.

This report presents a statistical analysis done on 6,065 rides from open dressage performance shows between 1 October thru 30 November 2007. This represents approximately 20% of all USDF Open Class scores in 2007. This analysis was undertaken to determine if there was a problem with performance, such that a “performance standard” needs to be implemented.

The statistical analysis concludes the following:

- Only .02 % (2 out of 3539) were in the 30-39% range (at Training and First Level)
- Only 1.78 % (158 total rides) were in the 40-49% range
- The data was “normal” (eg., bell curve) for all levels which makes it well suited for statistical analysis
- The ANOVA (Analysis of Variance) tool indicates that the average score was highest at Training level and decreased with increasing level. Below are the mean and standard deviation along with a pictorial representation of the confidence interval for the mean.

Level	N	Mean	StDev	
Training	1954	62.316	5.603	(--*)
1st	1585	61.739	5.221	(--*)
2nd	888	60.591	4.971	(--*--)
3rd	690	60.198	5.017	(---*--)
4th	453	61.313	4.824	(---*---)
PSG	274	60.339	4.713	(----*----)
Intermediare	157	60.122	5.201	(-----*-----)
GP Version B	64	59.116	4.751	(-----*-----)

Conclusions:

- Based on the detailed score analysis presented in the sections that follow, there does not seem to be evidence of poor riding at Training through Second Level. If poor riding was evident below 2nd Level, a large number of scores below 40% (insufficient) would have been expected.
- The Cumulative Distribution Function (CDF) analysis for 1st4, 2nd4, 3rd 3, 4th3 and PSG levels shows that putting a “hard” cut-off at 60% at any of those levels would

prevent approximately 40-60 percent of the rides from earning qualifying points. A cut-off of 55% would prevent 10-20 percent of the rides from earning points.

- Regression Analysis indicates no strong linkage exists between scores earned at 2-4 and future success at 3rd level; for 4th Level to PSG, the lowest scores earned at 4-3 are a better predictor of future PSG scores.
 - Without evidence of a valid predictive function, the implementation of qualifying scores must be questioned.

Observations & Recommendations

- The following key issues, as identified in the 2007 USDF Strategic Plan, should be considered in the context of this proposal:
 - The 2007 Strategic Plan emphasizes the need for the USDF to have strategic focus and that there is more work to be done for members than there are resources to accomplish it.
 - Financial issues facing members was identified as a major challenge for people involved in horses.
 - There is a declining trend in accessible training.
 - Finally, and most important to the qualifying rule question, the Strategic Plan recommends that the USDF should create a more “inclusive” culture.
- The authors of this analysis submit that devoting resources to implement and manage a “qualification rule” will dilute and detract people and financial resources that would be better directed towards furthering the USDF’s Strategic Plan.
 - Specifically, resources that would be consumed administering the qualification rule could be better used to further education and training.
- It is recommended that no qualification rule be implemented. This would satisfy the “more inclusive” goal of the strategic plan. Riders should utilize the show scores as guideposts for areas in which they need to improve.
- Rather than adding additional layers of complexity, there are mechanisms within current systems already in place to separate bad riding from good riding.
- The following existing systems can be used to address concerns regarding of “animal welfare” and / or accusations of animal cruelty:
 - Change the rider score to have a higher coefficient
 - If a qualification must be used, use the existing rider score as the qualifying mark, with a cut-off of 5.
 - Change the scoring scale such that the really poor rides are reflected by scores of 40% and below.
 - If the claim is that the poor riding is harmful to the horse, charge the TD to disqualify people whose riding so merits.
 - Charge the judges with disqualifying those riders who demonstrate poor riding.
 - Since judges fear being blacklisted for low scores or disqualification of riders, eliminate the ability for show managers to “judge shop” by implementing a “judging lottery” for judges.

Introduction

USEF Rule Change DR119.3, proposing a Performance Standard / Qualifying Score system for riders to move up the levels in competition, generated a great deal of interest and concern among riders. No clear problem statement or intended goal for the Rule Change was published or communicated to the USDF membership. However, it was indicated by Dressage Committee members that one reason for implementing Performance Standards was that judges feel they are seeing “30% rides”, with “confused and unhappy horses”. Furthermore, it was stated that the judges feel the implementation of a Qualifying Score system is an issue of horse welfare, as poor riding is a form of horse abuse. The implication is that a qualifying score system would improve the quality of riding in the show ring.

None of these statements were part of the rationale for the official proposed rule, but are identified in this analysis for background.

Three dominant questions and concerns emerged in the exchanges that occurred after the rule became widely known, which led the authors of this report to design four problem statements for examination:

1. What is the current state of dressage performances, as measured by test scores? Are there a significant number of low-scoring tests?
2. Is there a clear linkage between scores achieved at lower level tests and those later achieved at the next / higher levels? I.e., are scores at lower levels adequate / accurate predictors of future success at upper levels?
 - If there is no clear linkage between success at one level and success at higher levels, is there an alternate method that could be established to verify a rider’s readiness to advance?
3. If qualifying scores are mandated, whether for moving up the levels in competition or for entry to over-subscribed shows, what are the best ‘break points’ for determining which horse/rider pairs should move up/be permitted to enter? (Since all available data is framed in terms of horse/rider pairs, we will need to discuss the information in those terms.)

This paper examines published scores from USDF shows and provides several types of analysis and data visualizations to address the above questions / concerns.

This analysis produced four findings, which are discussed in detail on the following pages.

The graphs and charts produced by this analysis are contained in the remaining pages.

Summary of Findings, Conclusions, Concerns, and Recommendations

FINDING # 1:

There are almost no 30-39% rides and very few 40-49% rides occurring.

One-half of all rides achieve scores of 60% or more.

- Only .02 % (2 total rides) were in the 30-39% range
 - These scores occurred at Training and First Level; none occurred above First Level
- Only 1.78 % (158 total rides) were in the 40-49% range
 - These scores occurred at all levels.
- The mean score across each level was:

Training Level	62.316
First Level	61.739
Second Level	60.591
Third Level	60.198
Fourth Level	61.313
PSG	60.339
Intermed	60.122
GP	59.116

Conclusion:

If judges feel they are seeing “poor” rides, they are not documenting this on score sheets.

Concerns:

One possible reason given for the lack of low scores is that judges are worried they will not be hired for future jobs, will not be promoted, or may even be subjected to litigation if they are perceived as ‘low scoring’ judges.

With half of all rides scoring 60% or better, we must question the validity / utility / cost effectiveness of qualifying criteria applied to all riders in order to move up the levels in competition.

- A system of recording and reporting qualifying scores at one level before being allowed to compete at the next imposes a burden of cost and time on riders, show managers, and USDF staff.
 - These costs must be adequately projected and budgeted as part of implementing any such system.
- If a centralized system of recording these qualifying scores / points is developed, there must be provisions for dealing with the time delays that will occur between earning points and having the documentation available to the next show.

Recommendations:

Judges must be encouraged to use the full range of available scores and mark down poor rides. If all judges are consistent in scoring rides, there will be no outliers to be perceived as 'high' or 'low' scoring judges.

Show managers and riders must be better educated on the need / value of accurate judging.

FINDING # 2:

There is no clear linkage between good scores at lower level tests and good scores at higher level tests.

- Individual rider results were traced through competitions, as the rider moved up the levels with one or more horses
- While good scores at lower level tests did predict that a rider would make the move up the levels, the scores earned at the higher levels varied so widely that no clear patterns emerged which would indicate that a strong linkage exists between success at lower levels and success at higher ones.
- The pattern that did emerge was that low scores at higher level tests lead to higher wash-out rates at those levels; i.e., riders stopped entering those tests were they frequently earned poor scores
 - This is most evident in the declining numbers of rides as the levels increase.
- The score patterns also showed a high correlation to the horse being ridden: a rider frequently produced good 2nd Test 4 Scores on one horse and poor 3rd Level scores on a different horse, as well as vice-versa.
- For rides at 4th Lvl – PSG, the predictive results are more consistent WHEN THE LOWEST SCORES achieved by the same horse/rider pair are examined.
 - When the HIGHEST SCORES earned by a horse/rider pair are used, the correlation between success at 4th Lvl and success at PSG is not strong.

Conclusions:

Riders naturally self-regulate their showing: earning lower scores discourages repeated rides at that level or above and the number of riders competing at the higher levels tapers off as a result.

Scores at lower level tests are not strong predictors of success at higher levels.

Recommendations:

Given that qualifying criteria do not have such a strong predictive value as to justify their use for qualifying riders to move the levels in competition, other methods of judging rider readiness to advance should be considered.

In the interest of "inclusiveness" and education, there should be no qualifying requirements based on earning points at shows. It should be the role of the GMO's to foster inclusiveness in showing so people can learn how to improve: if people are not showing, they have no valid feedback on where they need to improve.

FINDING # 3

A stronger linkage exists between scores earned at 3rd Level Test 1 and future scores at other 3rd Level tests, than exists between 2nd Level Test 4 and future 3rd level tests.

A stronger linkage exists between the LOWEST scores achieved at 4th Level and scores achieved at PSG.

Recommendations:

A qualifying test for riders could be developed where the test ridden is more accurate predictor of the rider's skill and readiness for the next higher level.

- These could be administered by the GMOs as part of the schooling show regime.
- Riders could earn a stamp on their USDF membership card or a certificate at the show, which could become part of the entry package when they wish to compete at a recognized show.
- This would significantly reduce administrated costs to riders, show management, and USDF.
 - For advancement to 3rd level, the test ridden should be similar to 3-1.
 - For advancement to PSG, the test should be similar to PSG.

FINDING # 4

- The horse has an impact on the scores earned; the same rider can earn good scores on one horse and poor on another.
 - This situation does not allow sufficient insight into rider capabilities.
- Scoring is consistent across the Regions examined; no one Region scores consistently different from the others.
- Individual charts for each level are included in the body of this report. These show the statistical patterns for each level.

Recommendations:

Qualifying criteria have a usefulness for show managers and awards programs, where the goal is to select the best performances for inclusion in an over-subscribed show or to ensure awards are meaningful recognition of superior performance.

- If qualifying scores / points are mandated, the score ranges for earning points must be based on the scoring data.
- The first score at which points are awarded should be at the Median for each level.
 - Awarding points for performances at less than the Median, i.e., in the lower half of the Bell Curve, does not adequately identify and reward 'above average' performance.
 - The point award increments should probably be in 2 percentage point steps.
- These score ranges must be updated regularly to address the issue of score creep.
- The minimum scores selected for earning points should match up with all other awards programs in order to avoid confusion and/or duplication of efforts by riders to earn points for various awards/qualifications.

Furthermore, if a qualifying system is put in place, it must be for a trial period, with benchmarks established for re-evaluation of the impact on scores and show entries at the end of the trial.

- If an analysis of the scores indicates there has been no significant change in the overall pattern of scores earned or a decline in the numbers of show entries, the qualifying requirement should be withdrawn.

Scope of Data Used

The data used in this analysis comprises all scores recorded at all USDF shows held between 1 October – 30 November 2007. This data represents approximately 20% of all USDF shows / scores in 2007. The final data set selected for analysis included:

- 6,065 scores from Open Classes only.
 - Scores from Qualifying, Musical Freestyle, Regional Championship, JR/YR classes, Sport Horse Series, and any Restricted classes were eliminated from the analysis in order to ensure consistency in the data.
 - The GP Special scores also were not considered in the statistical analysis as they were from 1 set of judges and had high residuals.
 - FEI Young Horse tests were not used, since the scoring system does not fit the standard 100% scale and the direct equivalents to the Level Tests are not precise.
- Shows from 6 of the 7 Regions (Region 6 held no shows during the sample timeframe)
- **92** individual judges or judge pairs (treated as a single unit, since only one final score is reported)

Types of Analysis and Visualization Techniques Used

Summary:

The Competition data was analyzed to determine the presence of trends and score variability by region, level, and judge. The intent of the analysis was to validate the problem statements and determine, by way of score analysis, whether there was a statistically different proportion of poor riding at 3rd level.

Score data was examined by:

- Test levels
- Score ranges
- Within each Region
- Across all Regions

Analysis techniques used included:

- Regression and analysis of variance (ANOVA) tools to perform statistical tests.
- Data presentation in bar and graph charts

Discussion of Analysis Results

The data was “normally distributed,” e.g., the distribution of scores was “well behaved data” and was suitable for statistical analysis.

Analysis was done to see if any particular judges had significant effects on the scores.

- There were 92 judges represented, of which 49 (53%), judged 80% of the rides.
- No statistically significant effect on scores was found by judge.

Significant statistical variables:

- Region
- Level
- Number of judges at competition. Number of judges could be considered a proxy to estimate size of competition
- No individual judge was found to introduce a bias to scores
- Average scores decreased with increasing level. GP had the lowest average score
- There is statistically significant variability in scores by level
- There is statistically significant variability in scores by region
- There is statistically significant variability in scores by competition as measured by number of judges

Some regions typically produce higher scores than others. No reasons were attributed as to why that may be, though class sizes and numbers of shows, as well as access to quality instruction may be causative factors.

Score analysis:

- There were no scores below 38.667%
- There were no scores above 78%.
- There were only 2 scores in the 30-40% range, both in Region 1, both at 1st level, with 2 different riders.
- There were 141 scores between 40% - 50%
- There were 2227 scores between 50% - 60%
- There were 3599 scores between 60% - 70%
- There were 286 scores between 70% - 80% distributed as follows
 - 145 at Training level
 - 80 at First Level
 - 28 at Second Level
 - 14 at Third Level
 - 16 at Fourth Level
 - 3 at Prix St. George
 - 5 at Intermediare
 - 2 at Grand Prix

Specific Observations

Issue: Whether scores at lower level tests can be used as predictors of higher level success

The impression given is that the judges / sponsors of the rule change believe there is a correlation between successful scores at 2nd level and future success at 3rd level, as well as between scores at 4th level and future success at Prix St. Georges (PSG).

Observations: 2nd Level to 3rd Level

For the move from 2nd level to 3rd level, the data indicates the answer to be 'not so much'.

- Statistically, there was no significant difference between scores at 2nd Level Test Four and 3rd Level test 1:
 - The average score at 2-4 = 60.779, with a low of 47.381 and a high of 76.667
 - The average score at 3-1 = 60.012, with a low of 45.641 and a high of 71.795

Specifically, in the data available, there are 15 instances where a rider rode at both 2-4 and a 3rd level test. For this group:

- **3** riders rode the same horse at both levels and got better scores at 3rd than at 2-4.
 - The scores were (2-4 - 3rd): 59-60; 61-62; 59-61.
- **8** riders rode the same horse at both levels and got a poorer score at 3rd than at 2-4.
 - The scores were: 65-61; 58-51; 62-60; 62-61; 64-59; 55-49; 65-62; 60-53.
 - Notice that in this data set, only one rider did not meet any of the possible qualifying scores being suggested.
 - Four of these riders got 3rd lvl scores that were at or above the average for the level.
- **1** rider rode different horses at both levels and got a better score at 3rd.
 - Scores were: 63-64
- **3** riders rode different horses at both levels and got poorer scores at 3rd lvl.
 - Scores were: 69-67; 64-52; 76-64

What is interesting in this sample is that riders do give the impression of self-regulating the tests they enter: only one rider scored poorly at 2-4 (55%) and went on to ride 3-1 for a score of 49%.

The better predictor of success at Third Level tests appears to be the scores obtained at 3-1: riders who did poorly at 3-1 did poorly at other Third Level tests, with only a very loose correlation to the scores for 2-4.

- Riders scoring in the 50's at 3-1 went on to get scores in the 50% range on other 3rd level tests.
- Riders with at least one score in the 60's and one in the 50's at 3-1 had decidedly mixed results, with scores ranging from 45 - 72.
- Riders with all scores in the 60's at 3-1 produced scoring that was much more consistent: if the pair was scoring 63-64 at 3-1, then they earned scores similar scores at other 3rd lvl test.

Following up on this detailed analysis:

- There were 58 riders who rode 3-1 two or more times. Of these:
 - 17 had all their scores in the 50's
 - 19 had at least one score in the 60's and one in the 50's
 - 22 had all scores in the 60's
 - The lowest pairing were 50-49 and 56-49
 - The best scores were 63-64 and 68-63.

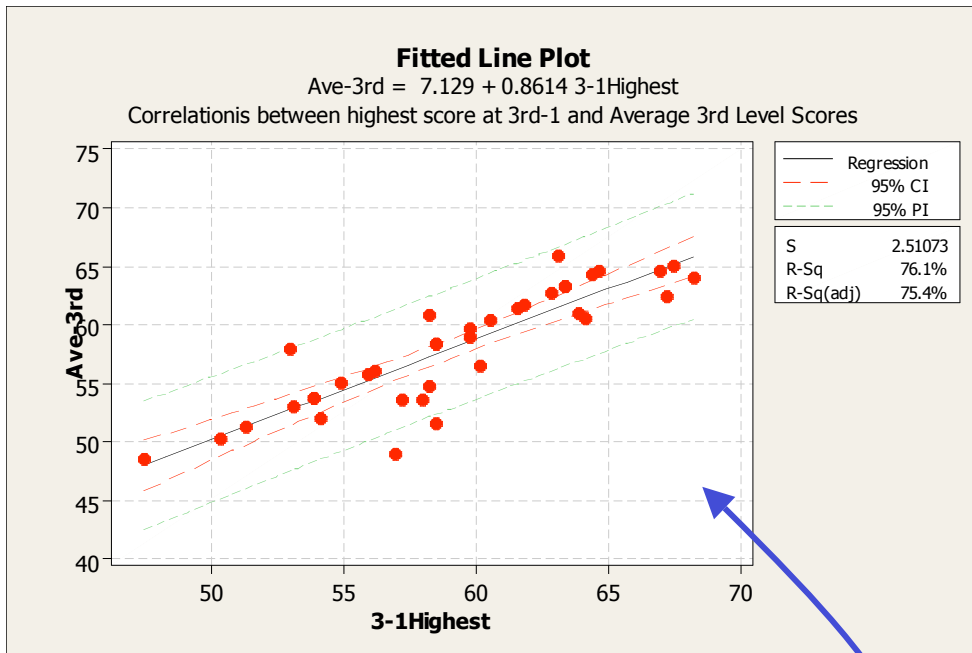
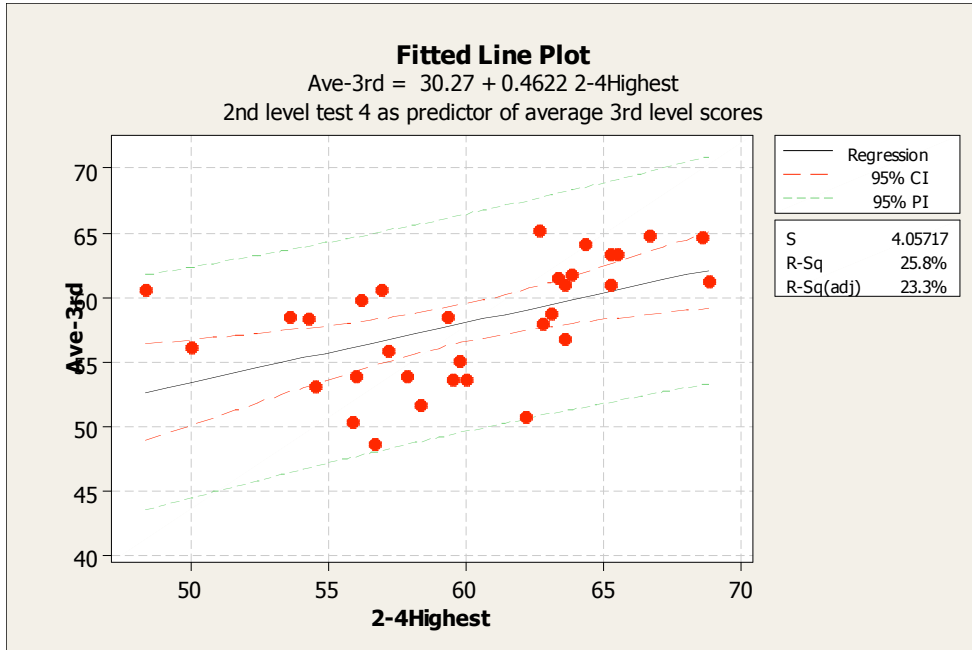
Once again, we see a nice distribution of the population - one third high, one third middling, one third low.

In looking at the scores with the largest variance between rides, the variable was the judge: these were the same horse/rider pairs, with 2 or more different judges.

Out of this group of 58, 30 went on to ride other 3rd level tests. Here is were a pattern truly emerges:

- Of the **17** with all their scores in the 50's, 9 went on to ride other 3rd level tests.
 - All of these scores were in the 50 percent range, with only one 60% reported.
- Of the **19** with at least one score in the 60's and one in the 50's, 11 rode additional 3rd lvl tests
 - With decidedly mixed results: scores ranged from 45 - 72.
- Of the **22** with all scores in the 60's, 10 went on to ride other 3rd lvl tests and the scoring was much more consistent
 - If the pair was scoring 63-64 at 3-1, then they earned scores similar scores at other 3rd lvl test.

Here, the data suggests that success at 3-1 is a much better predictor of success at other 3rd level tests.



Correlations: 2-4Highest, 3-1Highest, Ave-3rd

	2-4Highest	3-1Highest
3-1Highest	0.502	0.008
Ave-3rd	0.508	0.872
Cell Contents: Pearson correlation	0.003	0.000

87%
Correlation
coefficient

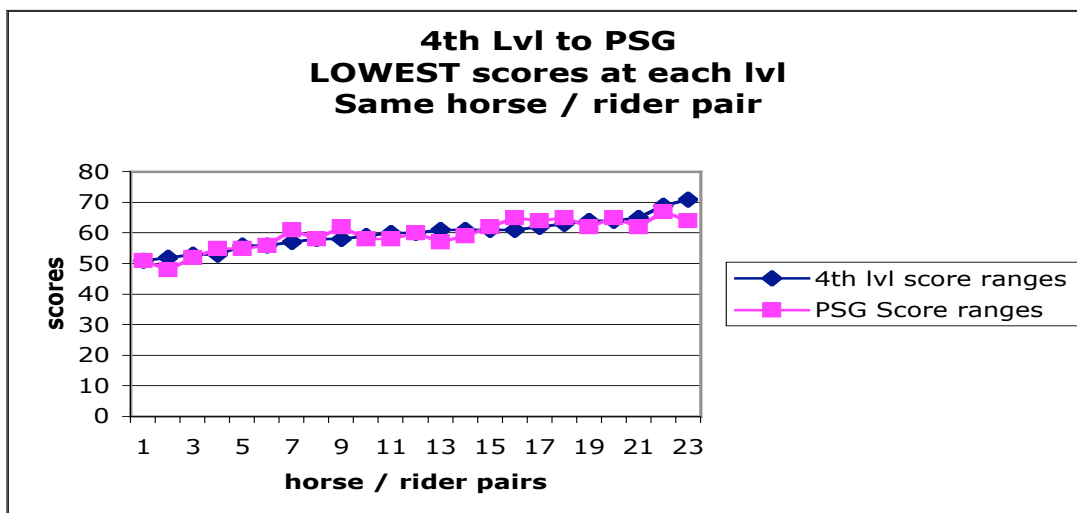
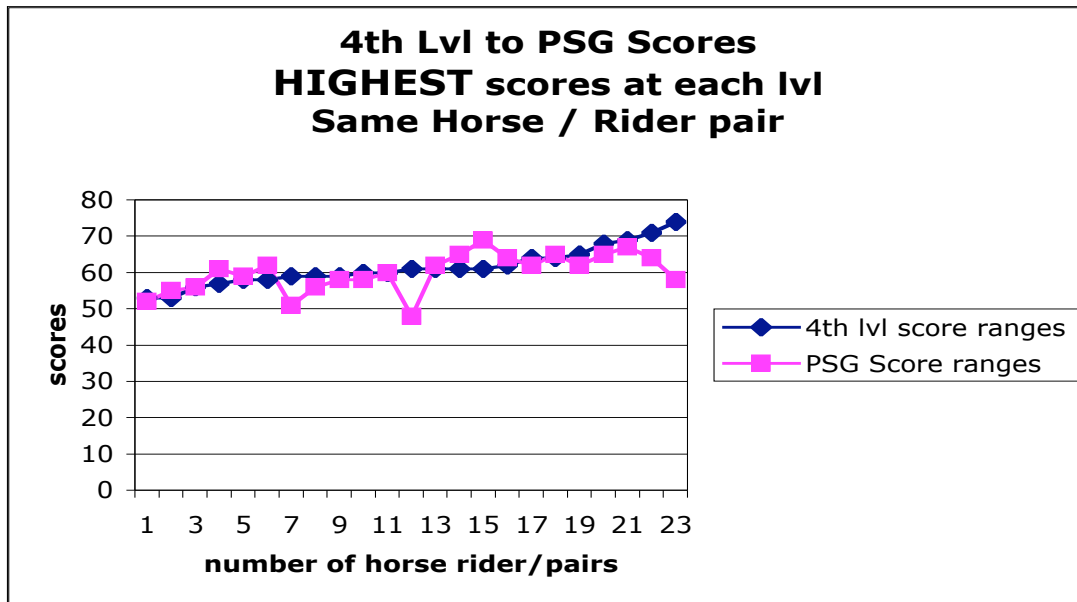
Observations: 4th Level to PSG

For the 4th Lvl to PSG transition, the data was examined in several ways:

- 23 cases where the same horse/rider pair rode at least one 4th Lvl and at least one PSG test
- 15 cases where the same rider rode two or more horses in 4th and PSG tests

The data was charted using different score pairings:

- Highest 4th against Highest PSG and
- Lowest 4th against Lowest PSG:



The highest scores a horse/rider pair achieved in any one 4th Lvl test were not strong predictors of success at PSG. However, the LOWEST scores they received did map more tightly to the PSG scores.

For the data where the same rider rode multiple horses at both levels, the data was too mixed to suggest any trends or correlations, other than that horse skill and training was likely to be of greater impact.

Data

Number of scores by Region and Level

Region	1st	2nd	3rd	4th	GP Special	GP Version B	Intermediare	PSG	Training	Grand Total	Percent By Region
1	437	278	194	158	5	20	38	80	389	1599	26%
2	83	34	22	22		2	3	12	184	362	6%
3	226	121	72	59	3	20	30	63	265	859	14%
4	125	29	32	25			4	14	125	354	6%
5	143	100	64	41		2	11	25	193	579	10%
7	323	191	163	97	4	12	47	56	504	1397	23%
8	57	36	46	12				1	44	196	3%
9	191	99	97	39		8	24	23	250	731	12%
Grand Total	1585	888	690	453	12	64	157	274	1954	6077	100%
Percent by Level	26%	15%	11%	7%	0%	1%	3%	5%	32%	100%	

The GP Special scores were not considered in the statistical analysis as they were from 1 set of judges. Only the 6065 scores from the other competitions were included in the detailed analysis.

Judge1 Number	Region								Grand Total	Percent of Scored	Cum Percent
	1	2	3	4	5	7	8	9			
Grand Total	1599	362	859	354	579	1397	196	731	6077	100%	
66					82	122			204	3.36%	3.36%
36	48					100		39	187	3.08%	6.43%
40			57	62	64				183	3.01%	9.45%
60	179								179	2.95%	12.39%
73					65	105			170	2.80%	15.19%
55	53					101			154	2.53%	17.72%
2			48					100	148	2.44%	20.16%
8	70		70						140	2.30%	22.46%
31	44					70		19	133	2.19%	24.65%
22		17	69				44		130	2.14%	26.79%
13	127								127	2.09%	28.88%
5			61			24	38		123	2.02%	30.90%
81	66		51						117	1.93%	32.83%
75	114								114	1.88%	34.70%
62		110							110	1.81%	36.51%
74						108			108	1.78%	38.29%
16		41			62				103	1.69%	39.99%
54								103	103	1.69%	41.68%
14								102	102	1.68%	43.36%
34	97								97	1.60%	44.96%
37			59			38			97	1.60%	46.55%
17					63	28			91	1.50%	48.05%
80			91						91	1.50%	49.55%
92								86	86	1.42%	50.96%
76						85			85	1.40%	52.36%
85	84								84	1.38%	53.74%
50	82							1	83	1.37%	55.11%
6		33	49						82	1.35%	56.46%
10						21		60	81	1.33%	57.79%
44						80			80	1.32%	59.11%
79								80	80	1.32%	60.42%
87			52	26					78	1.28%	61.71%
15					77				77	1.27%	62.98%
30						77			77	1.27%	64.24%
90				33		43			76	1.25%	65.49%
47	74								74	1.22%	66.71%
19							70		70	1.15%	67.86%
49					70				70	1.15%	69.01%
61								69	69	1.14%	70.15%
27		68							68	1.12%	71.27%
3				67					67	1.10%	72.37%
89		8	59						67	1.10%	73.47%
41						63			63	1.04%	74.51%
71	60								60	0.99%	75.50%
32	30					29			59	0.97%	76.47%
29					58				58	0.95%	77.42%
45	58								58	0.95%	78.38%
63				37	18				55	0.91%	79.28%
52				52					52	0.86%	80.14%
70			50						50	0.82%	80.96%
77	50								50	0.82%	81.78%
83	39		9						48	0.79%	82.57%
42	46								46	0.76%	83.33%
24	19		26						45	0.74%	84.07%
11				44					44	0.72%	84.80%
56							44		44	0.72%	85.52%
18	42								42	0.69%	86.21%
26			42						42	0.69%	86.90%
53	22							19	41	0.67%	87.58%
67						41			41	0.67%	88.25%
21		39							39	0.64%	88.89%
65						39			39	0.64%	89.53%
33								37	37	0.61%	90.14%
88	37								37	0.61%	90.75%
12						36			36	0.59%	91.34%
72	36								36	0.59%	91.94%
28	34								34	0.56%	92.50%
43	34								34	0.56%	93.06%
91						34			34	0.56%	93.62%
82				33					33	0.54%	94.16%

Summary of
scores by Region
and judge

Analysis of Variance by Judge
One-way ANOVA: Score versus Judge1Number

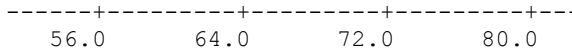
Source	DF	SS	MS	F	P
Judge1Number	91	16911.5	185.8	7.24	0.000
Error	5973	153368.5	25.7		
Total	6064	170280.0			

S = 5.067 R-Sq = 9.93% R-Sq(adj) = 8.56%

Individual 95% CIs For Mean Based on Pooled StDev

Judge	N	Mean	StDev	CI
1	16	59.310	5.155	(--*--)
2	148	60.263	6.180	(*)
3	67	58.642	5.214	(*--)
4	2	58.563	4.861	(-----*-----)
5	122	63.805	4.552	(*)
6	82	59.483	5.806	(*--)
7	17	61.718	3.732	(--*--)
8	140	62.687	5.113	(*)
9	16	58.374	4.724	(--*--)
10	81	62.108	5.960	(--*)
11	44	61.109	4.834	(*--)
12	36	62.634	2.937	(--*)
13	127	62.116	4.484	(*)
14	102	60.934	5.003	(*)
15	77	62.776	5.767	(*--)
16	103	61.900	4.921	(*--)
17	91	61.272	5.478	(--*)
18	42	60.648	4.655	(--*)
19	70	61.733	3.681	(*--)
20	25	63.072	5.447	(--*--)
21	39	64.703	3.958	(--*)
22	130	61.757	3.969	(*)
23	1	73.143	*	(-----*-----)
24	45	58.561	5.549	(--*)
25	22	60.561	5.393	(--*--)
26	40	61.918	6.943	(--*)
27	68	62.484	4.833	(*--)
28	34	58.072	3.521	(--*--)
29	58	62.287	6.427	(--*)
30	76	61.622	6.829	(*)
31	133	60.497	5.278	(*)
32	58	61.608	4.523	(--*)
33	37	56.884	5.410	(--*)
34	97	60.606	6.532	(--*)
35	15	65.874	3.288	(--*----)
36	187	59.712	6.023	(*)
37	94	61.800	5.032	(*--)
38	8	63.899	6.434	(----*----)
39	6	60.438	4.136	(-----*-----)
40	183	61.195	4.835	(*)
41	63	64.451	5.102	(--*)
42	46	58.217	5.938	(--*)
43	34	61.334	5.572	(--*)
44	80	64.974	3.288	(*--)
45	58	61.685	5.479	(--*)
46	16	61.377	5.170	(--*--)
47	74	60.042	4.706	(*)
48	3	61.542	2.444	(-----*-----)
49	70	62.735	3.775	(*--)

50	83	64.316	6.176	(*-)
51	16	62.227	6.140	(--*--)
52	52	58.110	7.060	(-*)
53	40	58.987	5.415	(-*-)
54	103	61.093	5.373	(*-)
55	154	61.117	4.831	(*)
56	44	57.327	6.821	(-*-)
57	26	59.030	5.512	(--*--)
58	20	58.560	5.228	(--*--)
59	17	63.120	4.845	(--*--)
60	177	60.160	5.286	(*)
61	69	60.404	3.342	(-*)
62	110	60.730	4.447	(*)
63	55	60.650	4.266	(-*)
64	22	63.063	5.948	(--*--)
65	39	63.427	4.556	(-*-)
66	204	62.320	3.363	(*)
67	41	63.709	4.017	(-*-)
68	12	61.398	4.621	(--*--)
69	31	61.636	4.678	(-*-)
70	50	64.473	5.163	(-*)
71	60	63.258	3.697	(-*-)
72	36	60.698	4.275	(-*-)
73	170	60.221	4.481	(*)
74	108	63.627	6.231	(-*)
75	114	64.431	4.759	(-*)
76	85	61.308	4.141	(-*)
77	50	59.117	8.130	(-*-)
78	28	59.045	4.614	(--*--)
79	80	60.918	4.006	(*-)
80	91	60.765	5.503	(*)
81	117	61.287	4.569	(-*)
82	33	58.963	4.801	(-*-)
83	48	61.218	4.315	(-*)
84	6	62.302	4.964	(----*----)
85	84	62.308	3.793	(*)
86	29	62.630	3.465	(-*-)
87	78	60.721	3.907	(-*)
88	37	60.375	5.952	(-*-)
89	67	66.054	5.531	(-*)
90	76	60.859	4.909	(*)
91	34	60.656	3.520	(-*-)
92	86	60.349	5.394	(*-)



Regression Analysis: Score versus Region, Level, Judge & Number of Judges

The regression equation is
 Score = 62.1 - 0.0808 Region - 0.435 LevelNumber + 0.00131 Judge1Number
 + 0.308 CountJudges

Predictor	Coef	SE Coef	T	P	VIF
Constant	62.0744	0.2543	244.08	0.000	
Region	-0.08084	0.02402	-3.36	0.001	1.045
LevelNumber	-0.43546	0.04009	-10.86	0.000	1.035
Judge1Number	0.001313	0.002551	0.51	0.607	1.018
CountJudges	0.3081	0.1599	1.93	0.054	1.085

S = 5.24690 R-Sq = 2.0% R-Sq(adj) = 2.0%

Analysis of Variance

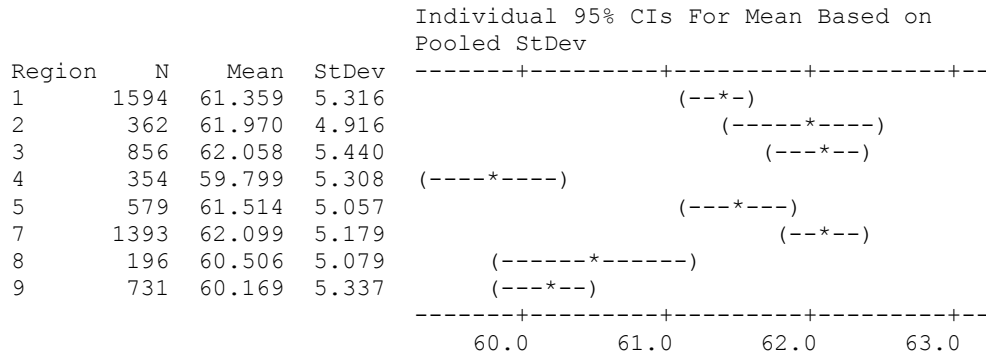
Source	DF	SS	MS	F	P
Regression	4	3448.23	862.06	31.31	0.000
Residual Error	6060	166831.81	27.53		
Total	6064	170280.04			

Source	DF	Seq SS
Region	1	185.65
LevelNumber	1	3157.98
Judge1Number	1	2.45
CountJudges	1	102.15

One-way ANOVA: Score versus Region

Source	DF	SS	MS	F	P
Region	7	3349.2	478.5	17.36	0.000
Error	6057	166930.8	27.6		
Total	6064	170280.0			

S = 5.250 R-Sq = 1.97% R-Sq(adj) = 1.85%

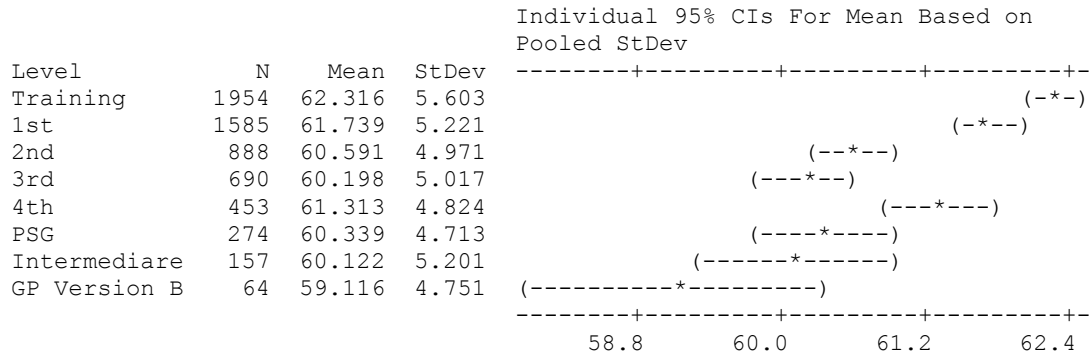


Pooled StDev = 5.250

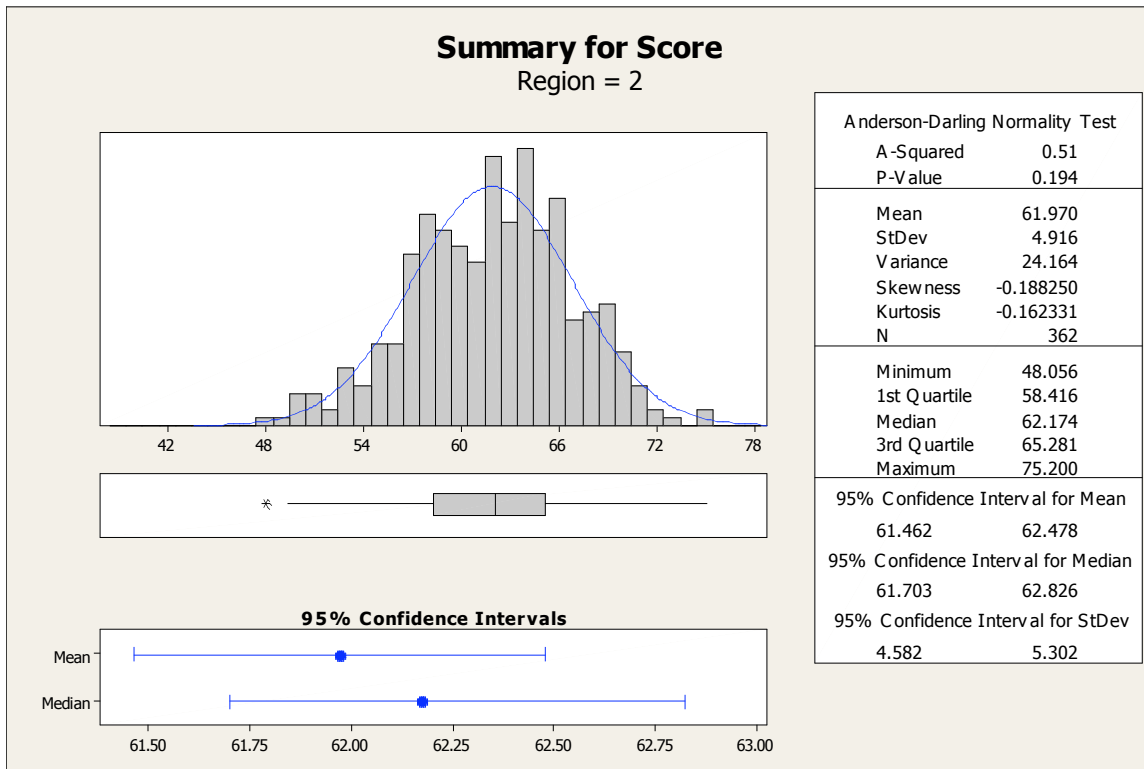
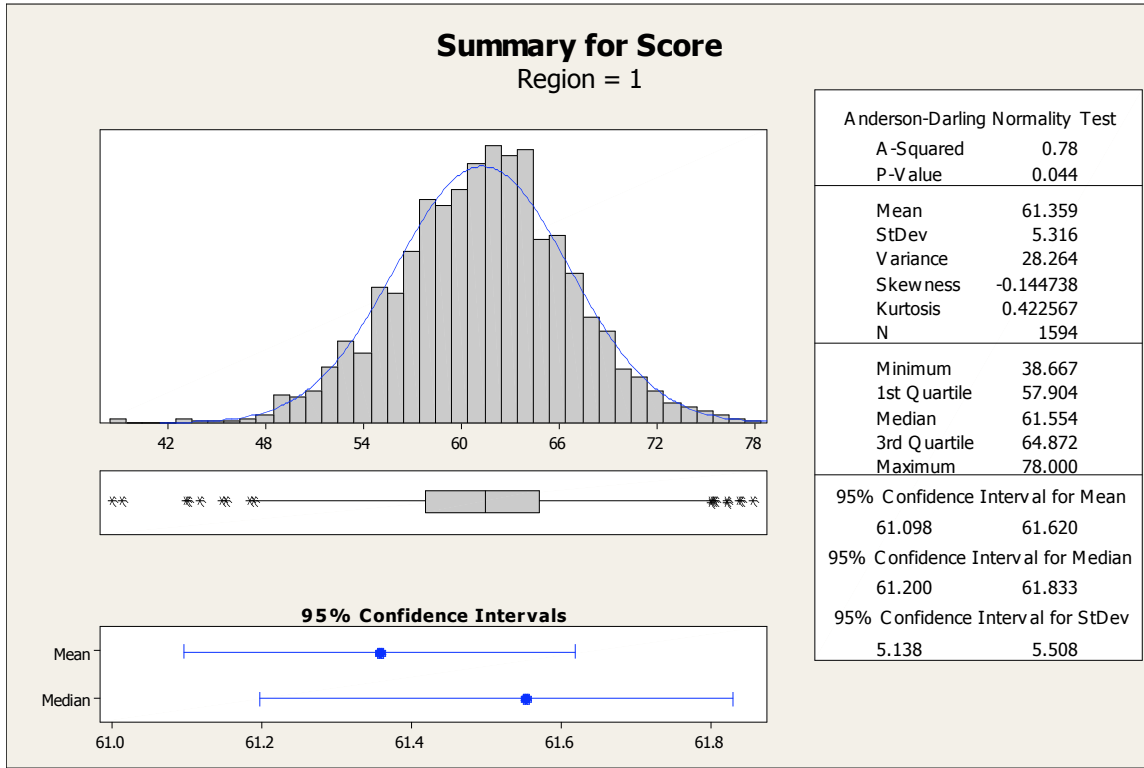
One-way ANOVA: Score versus LevelName

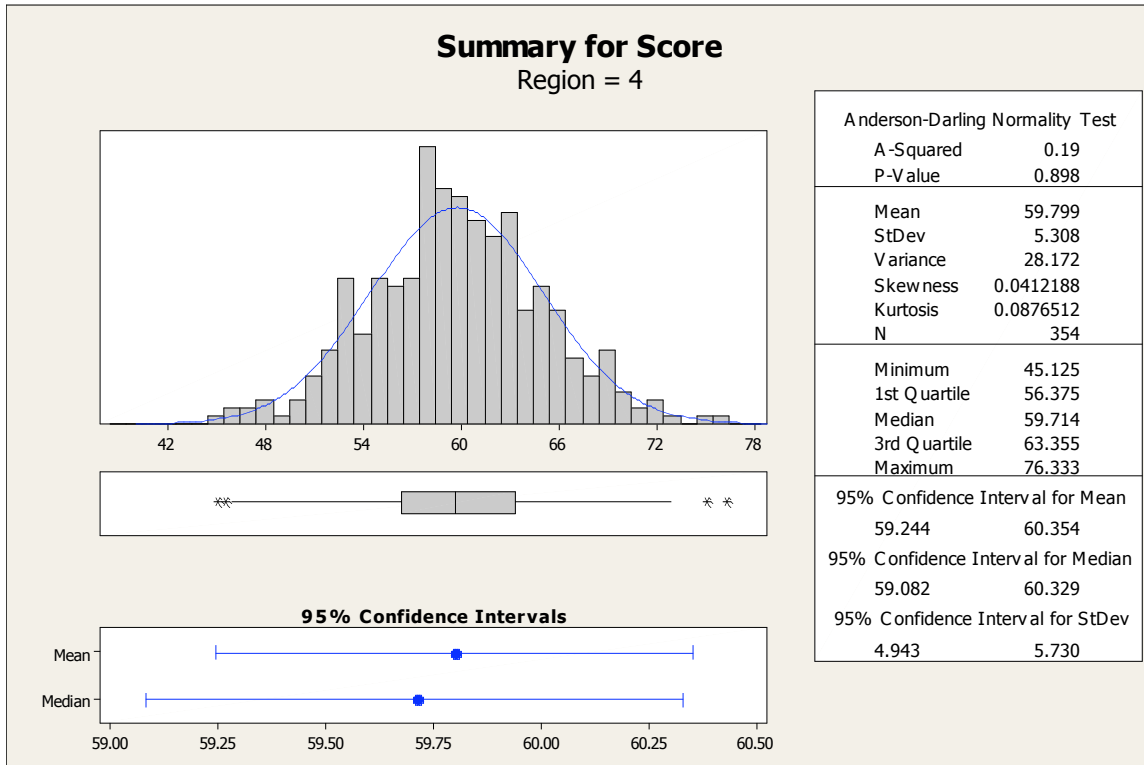
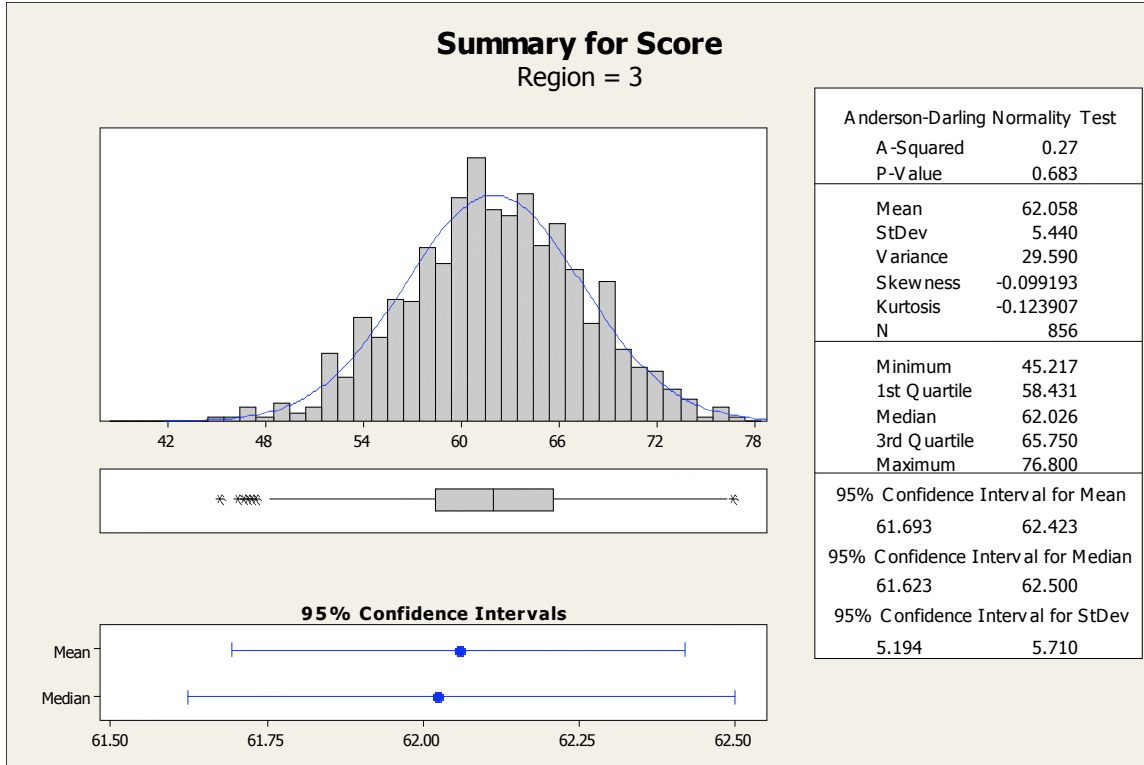
Source	DF	SS	MS	F	P
LevelName	7	4300.3	614.3	22.42	0.000
Error	6057	165979.7	27.4		
Total	6064	170280.0			

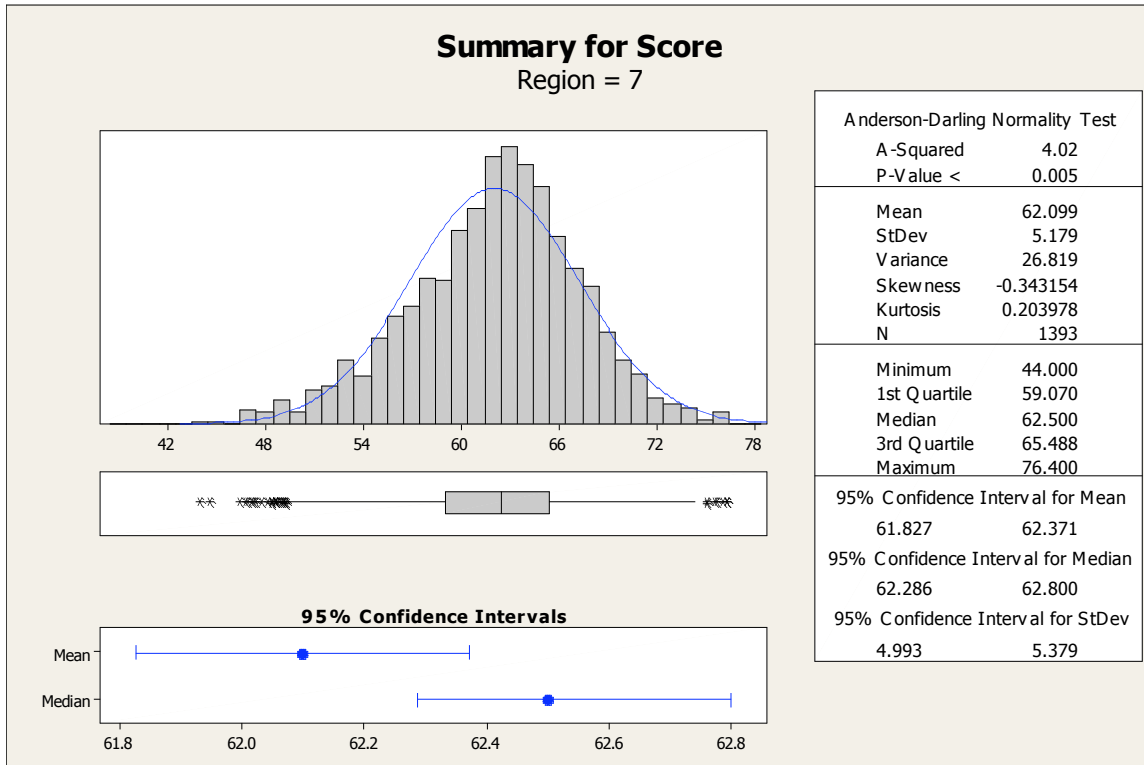
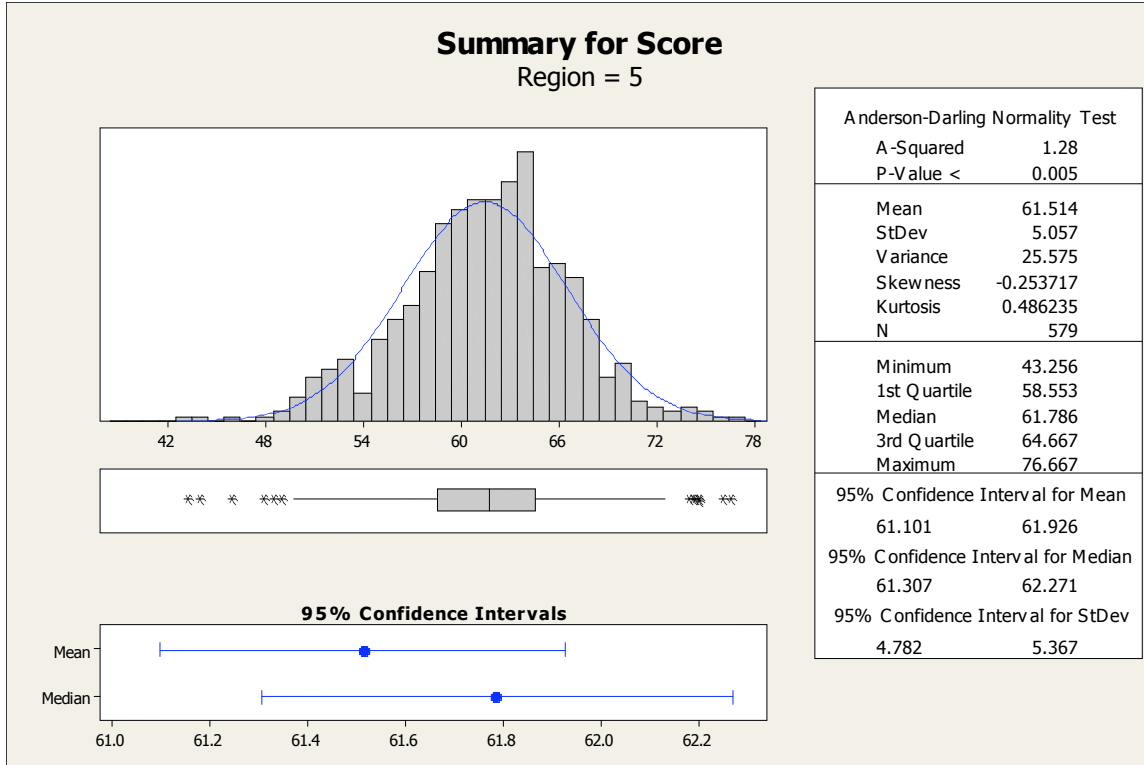
S = 5.235 R-Sq = 2.53% R-Sq(adj) = 2.41%

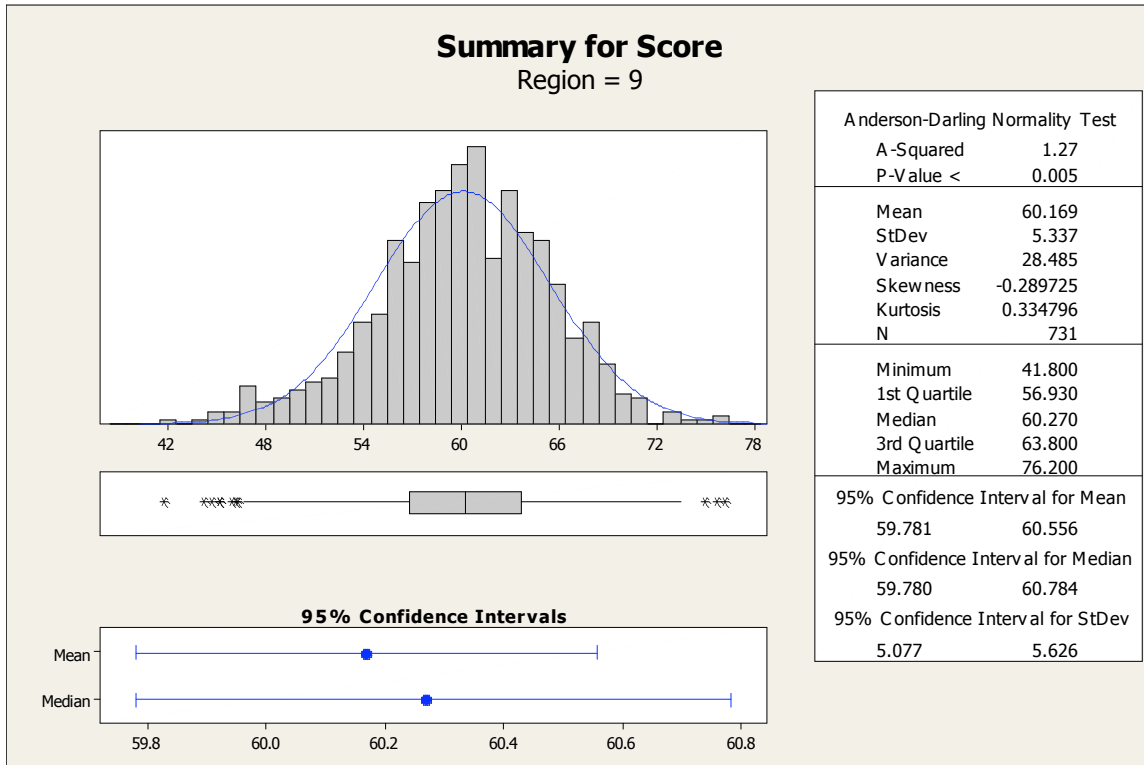
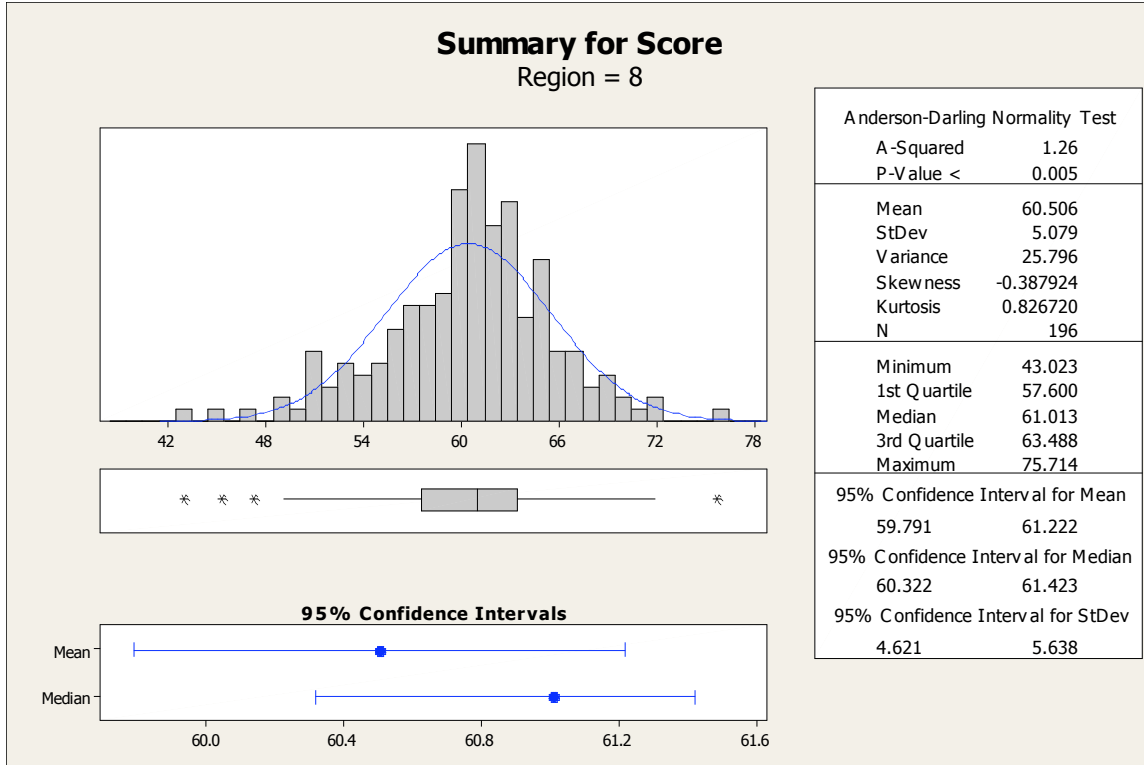


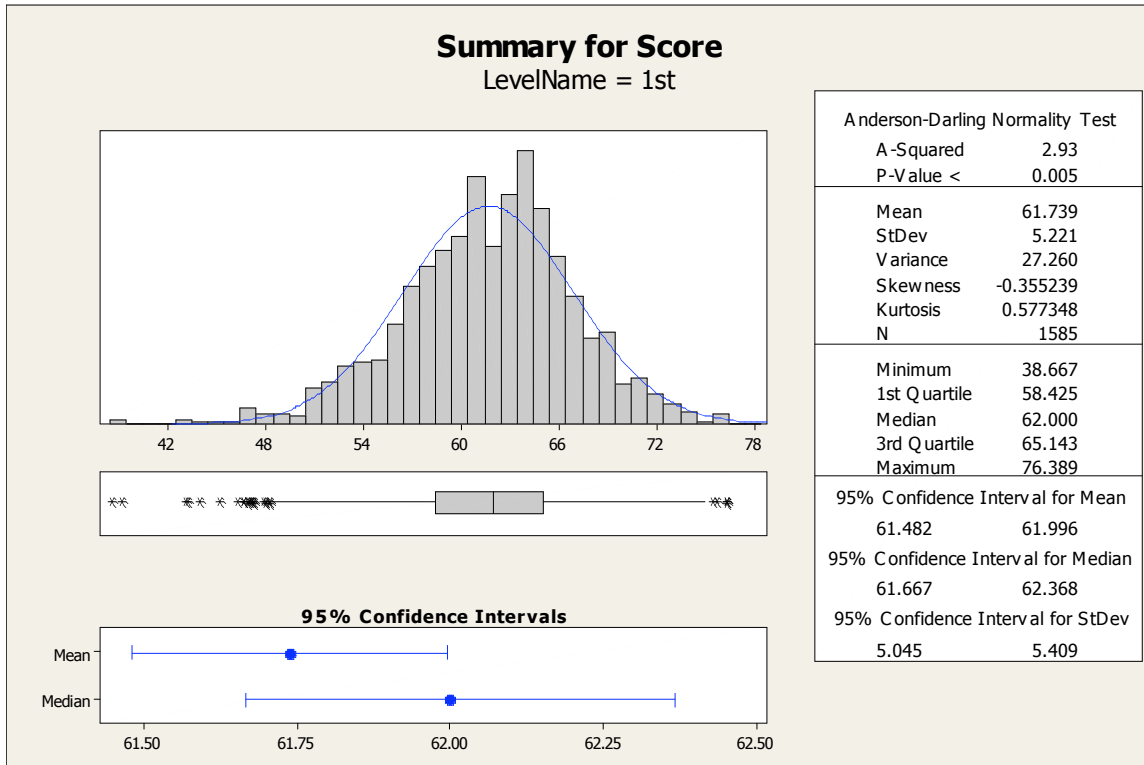
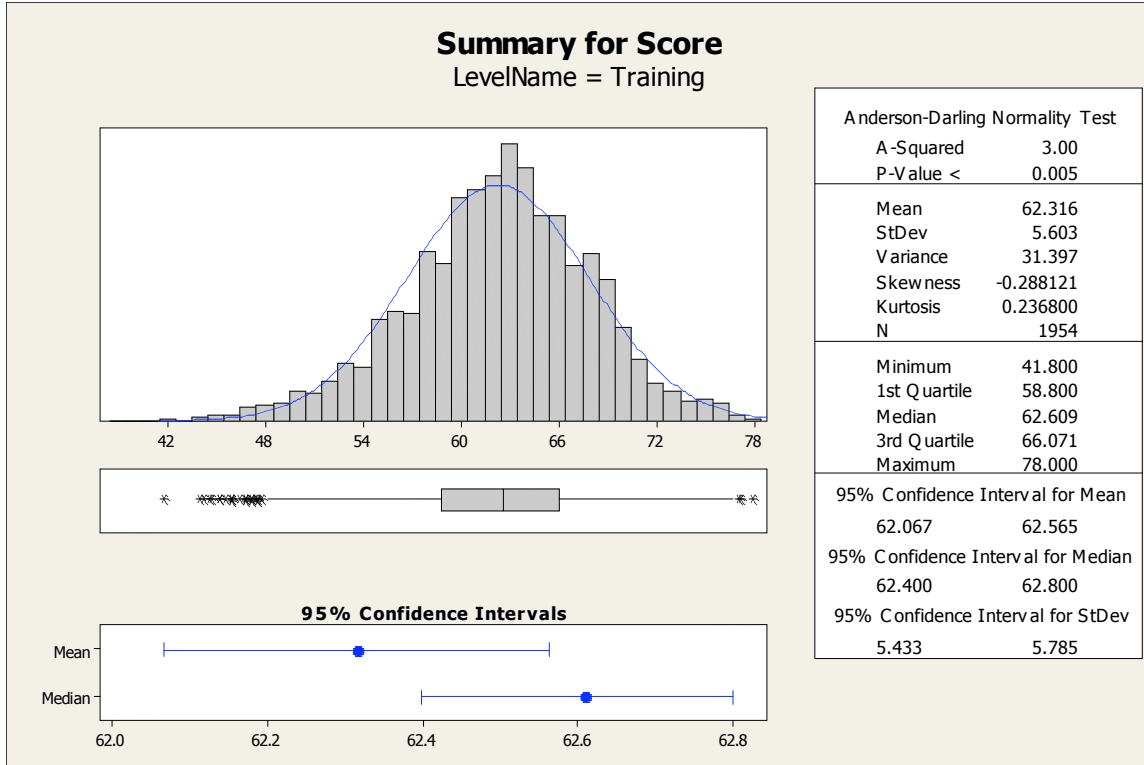
Pooled StDev = 5.235

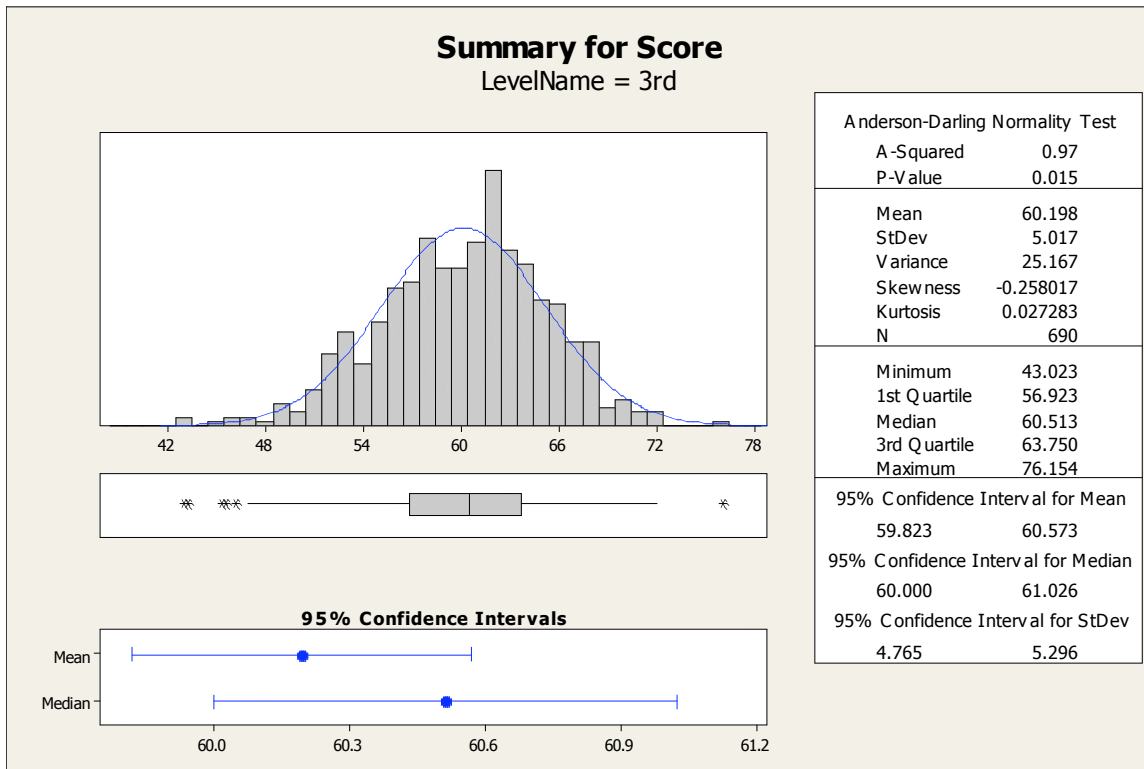
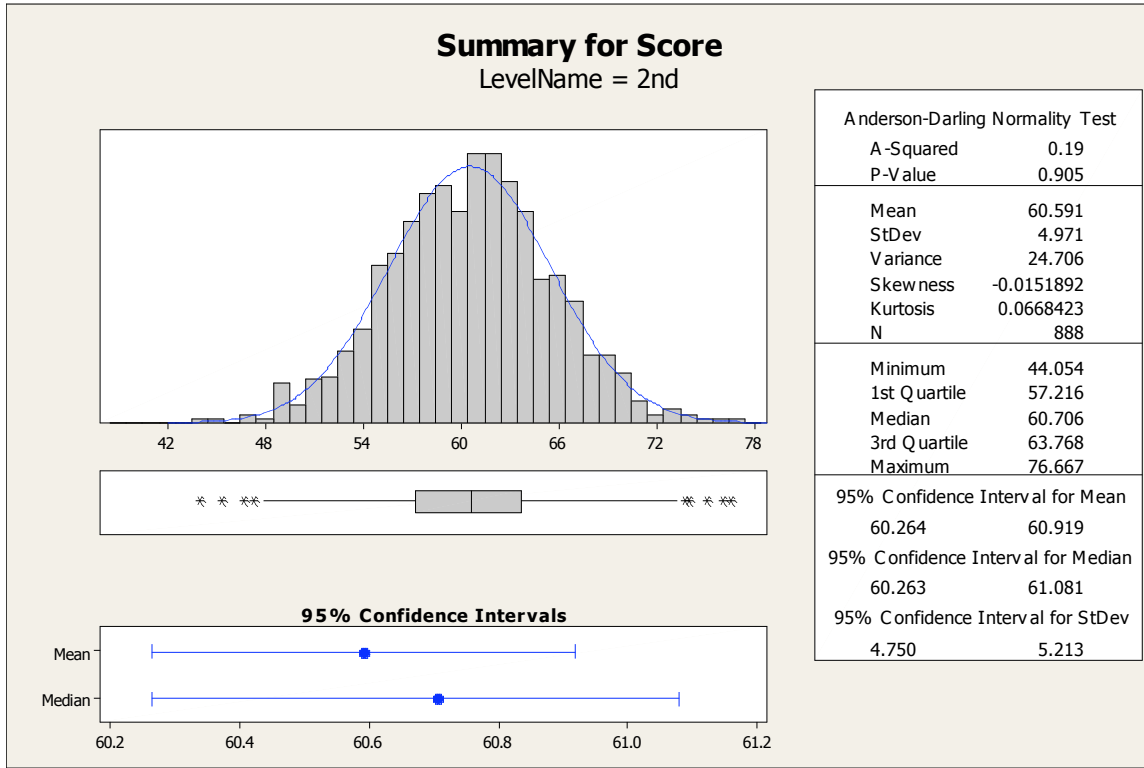


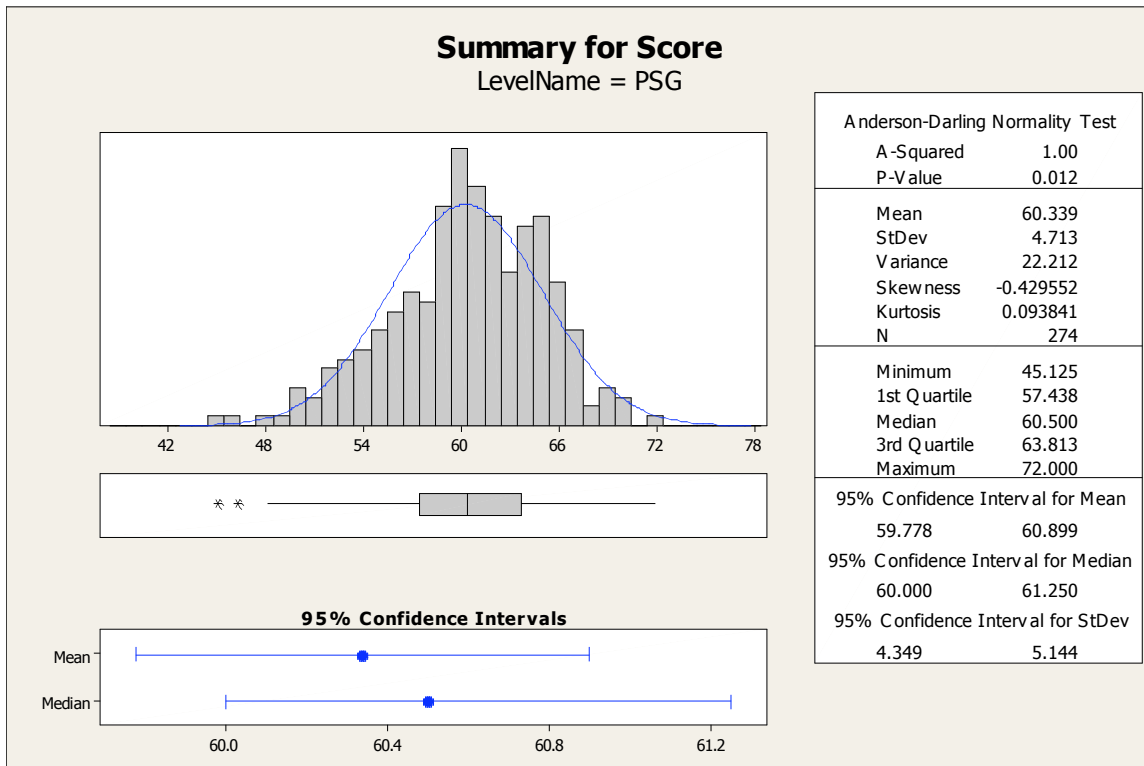
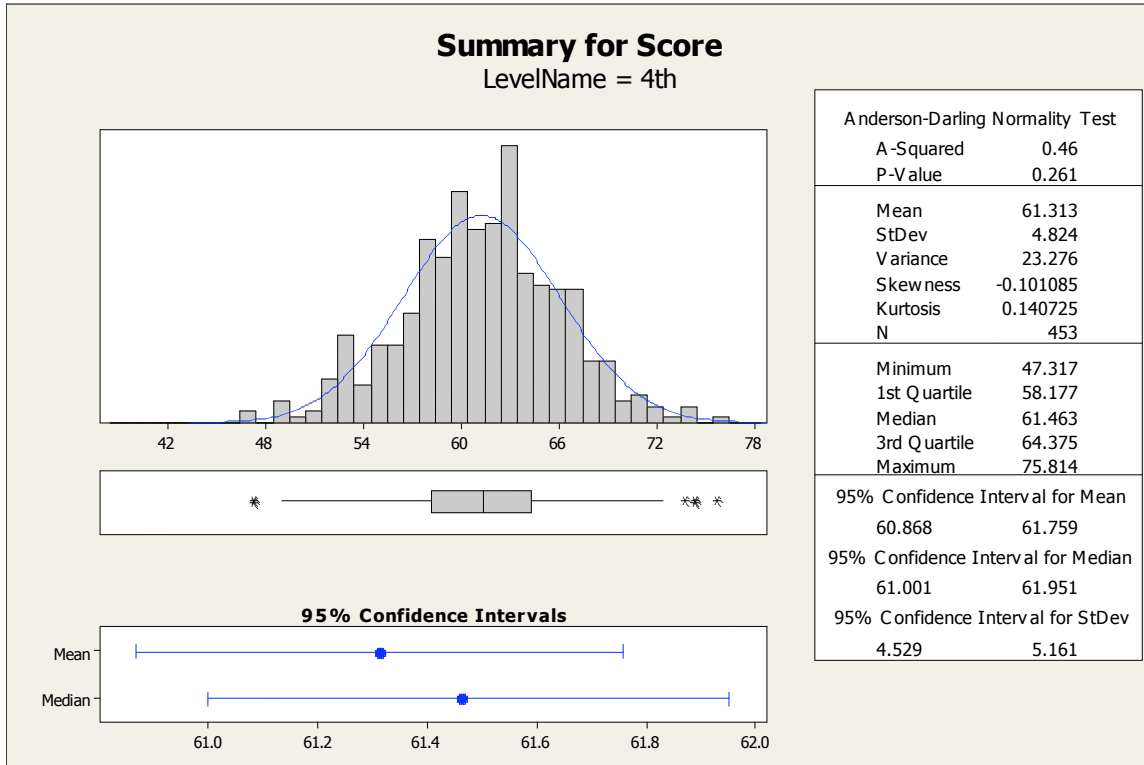


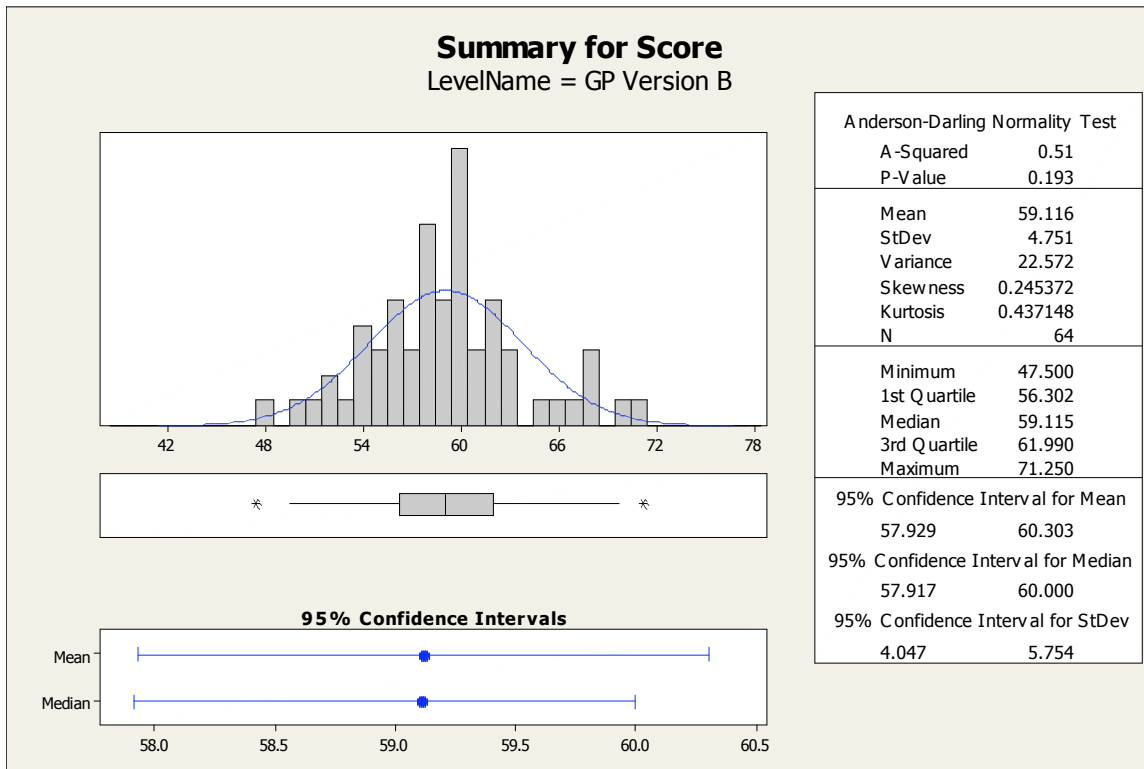
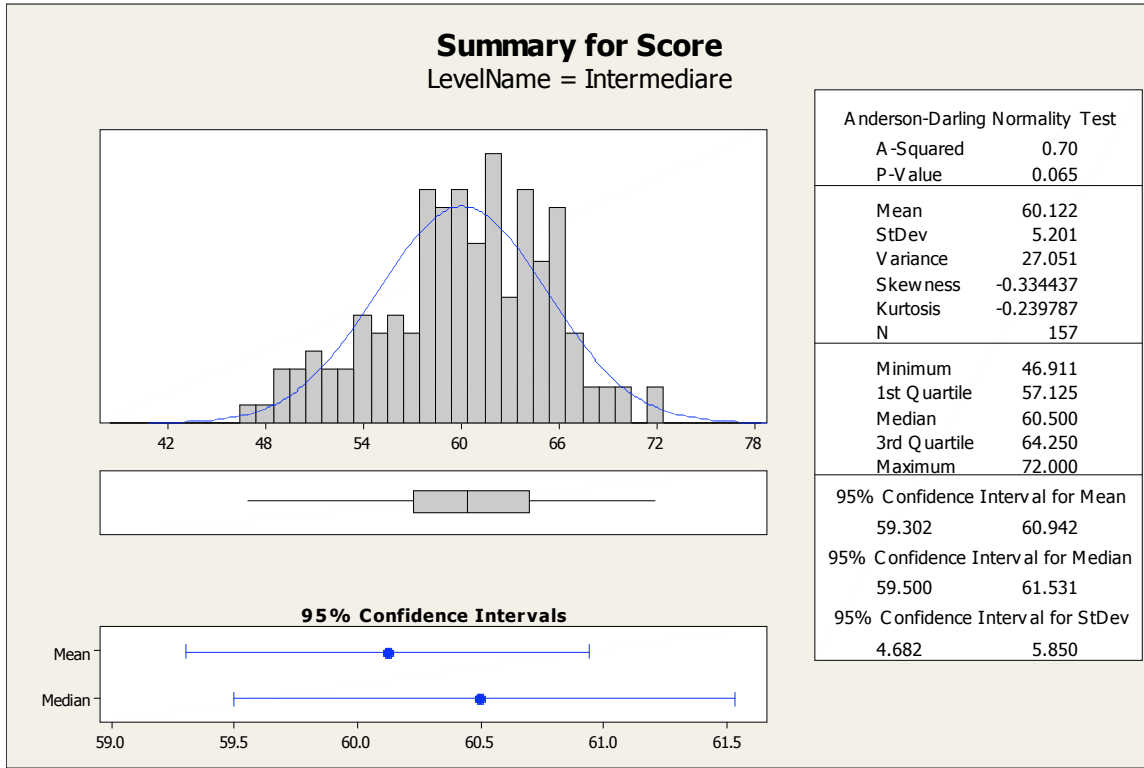




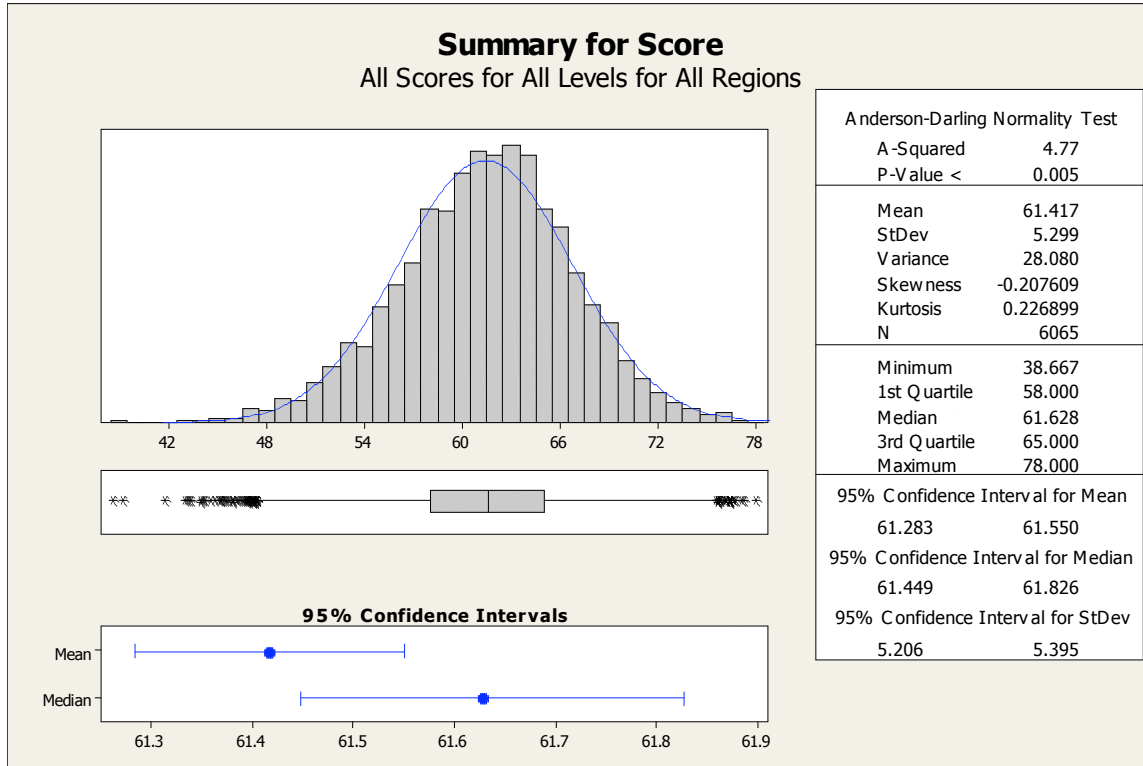


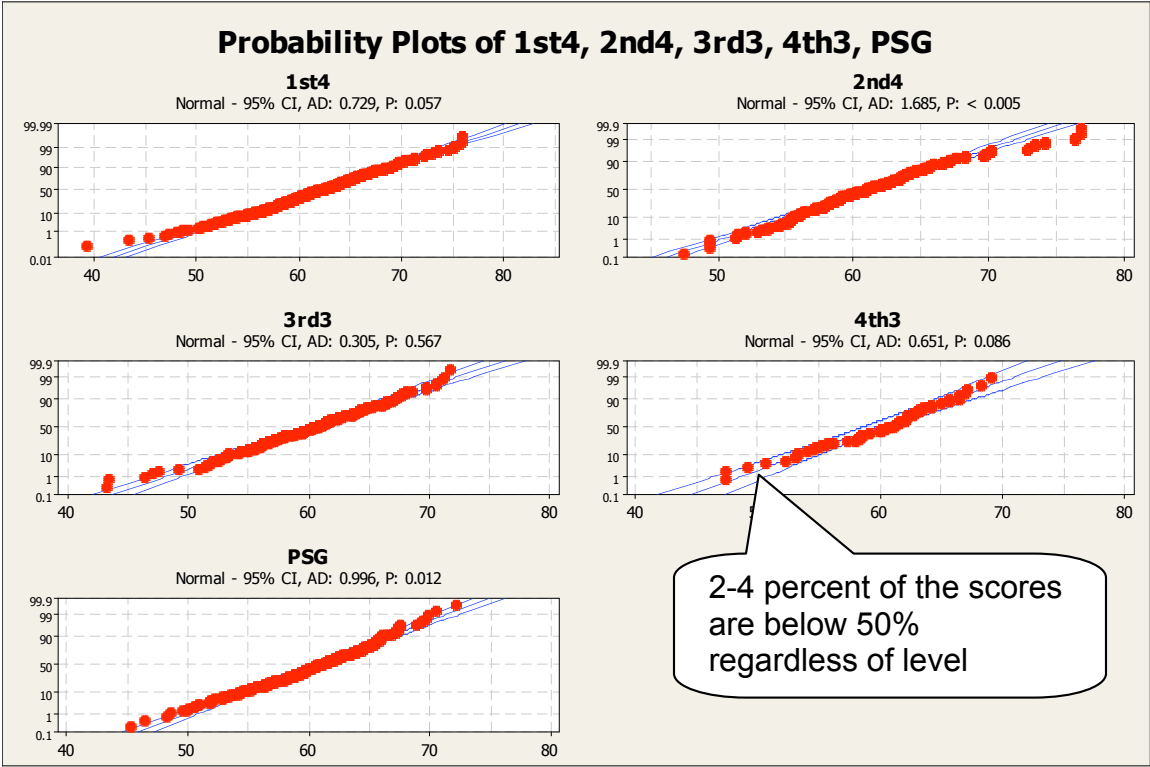
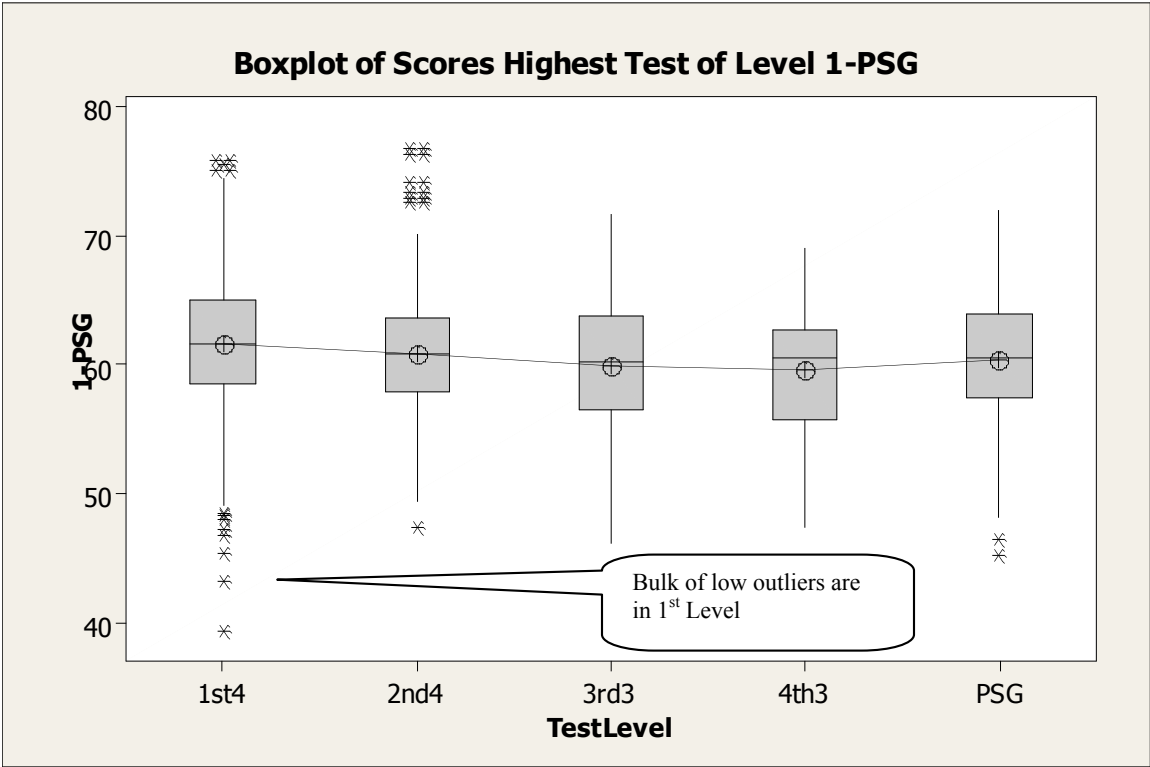


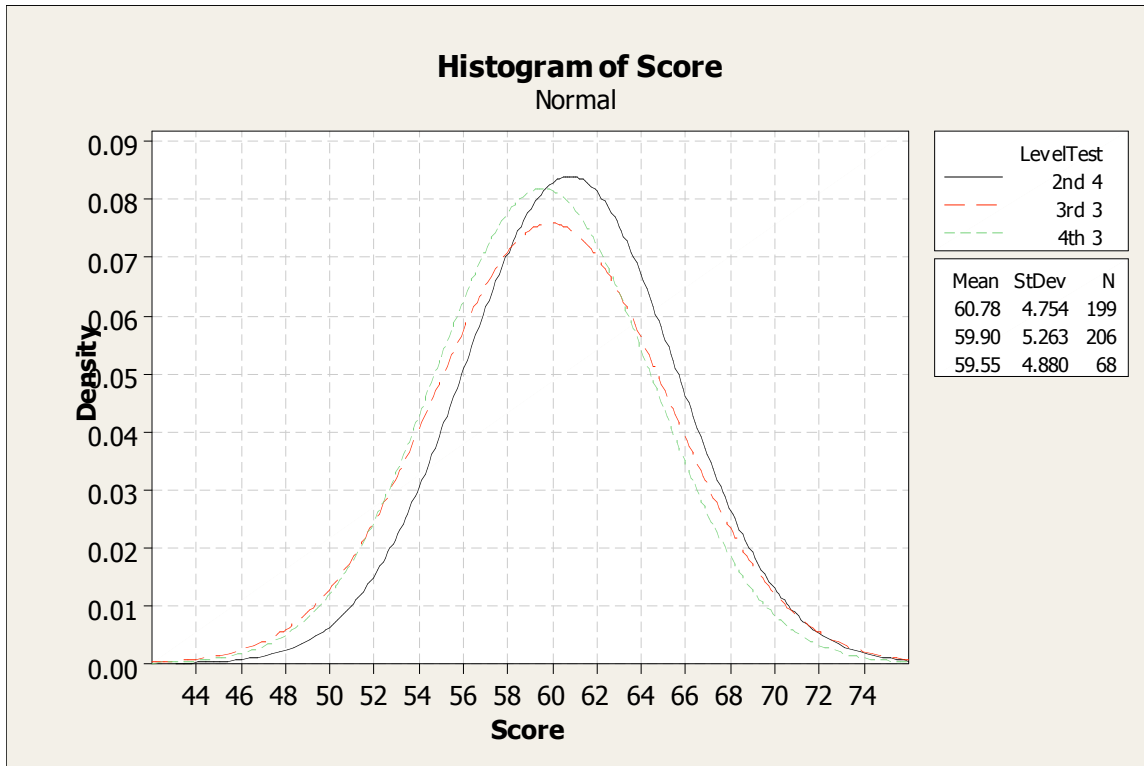
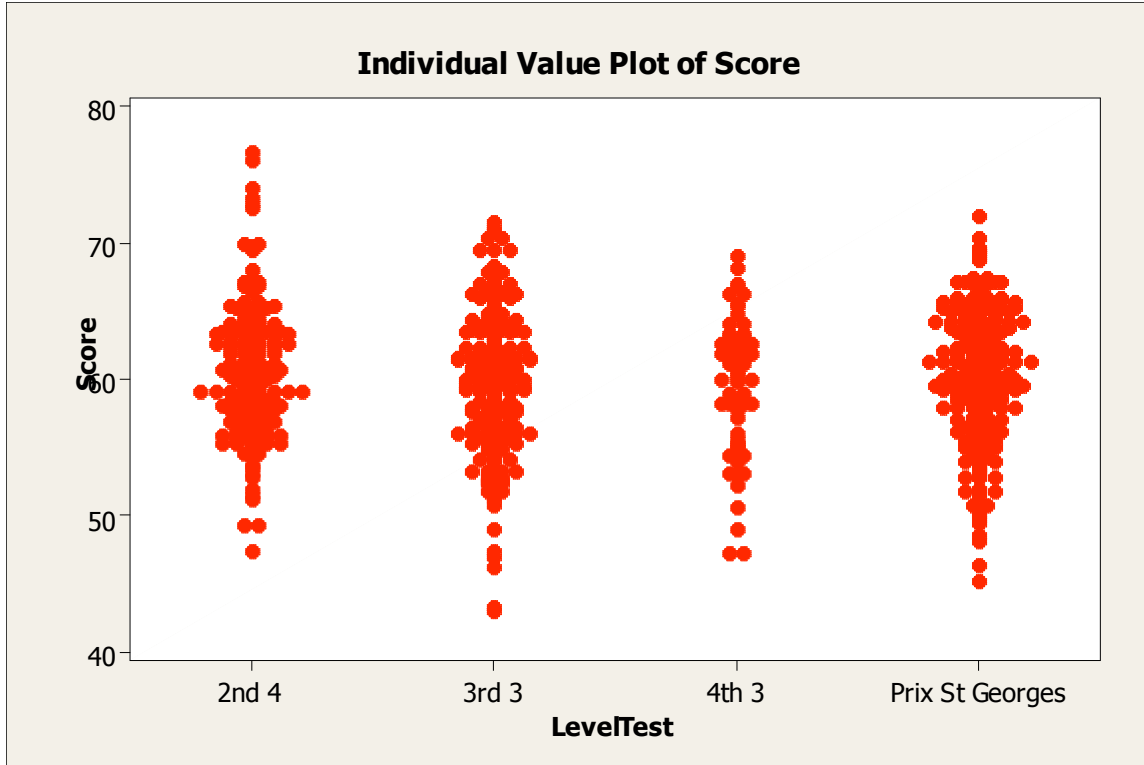


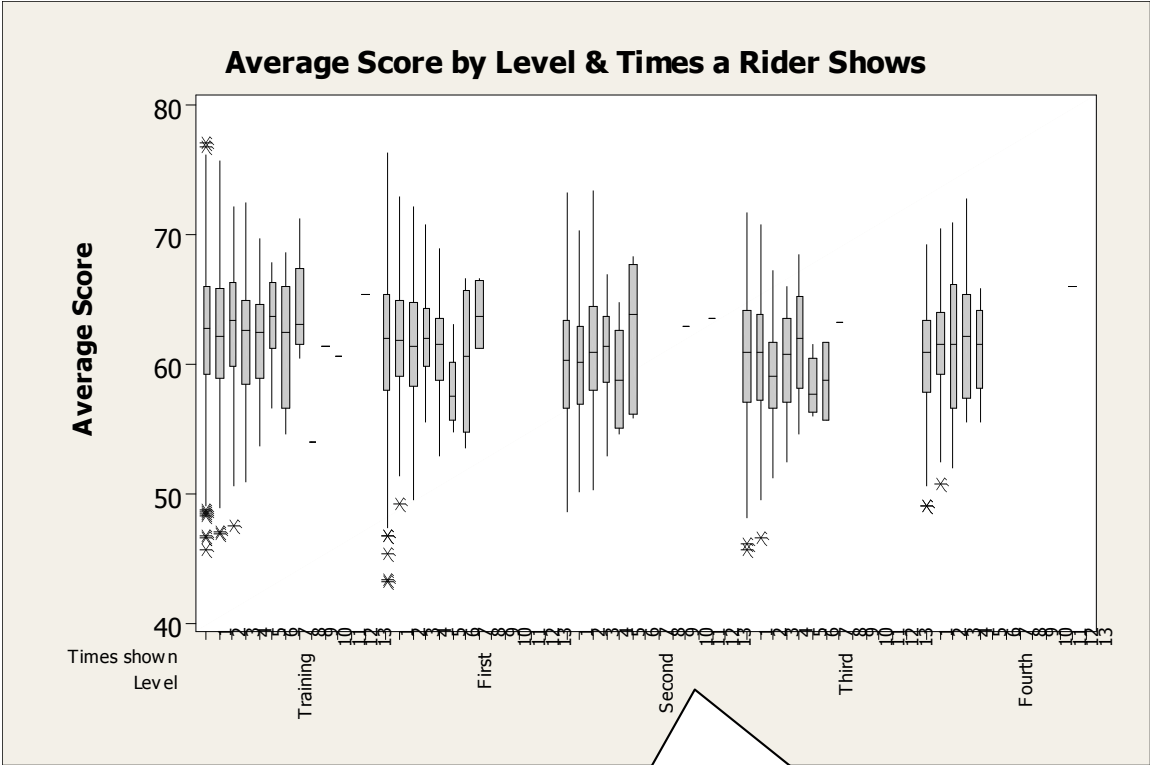
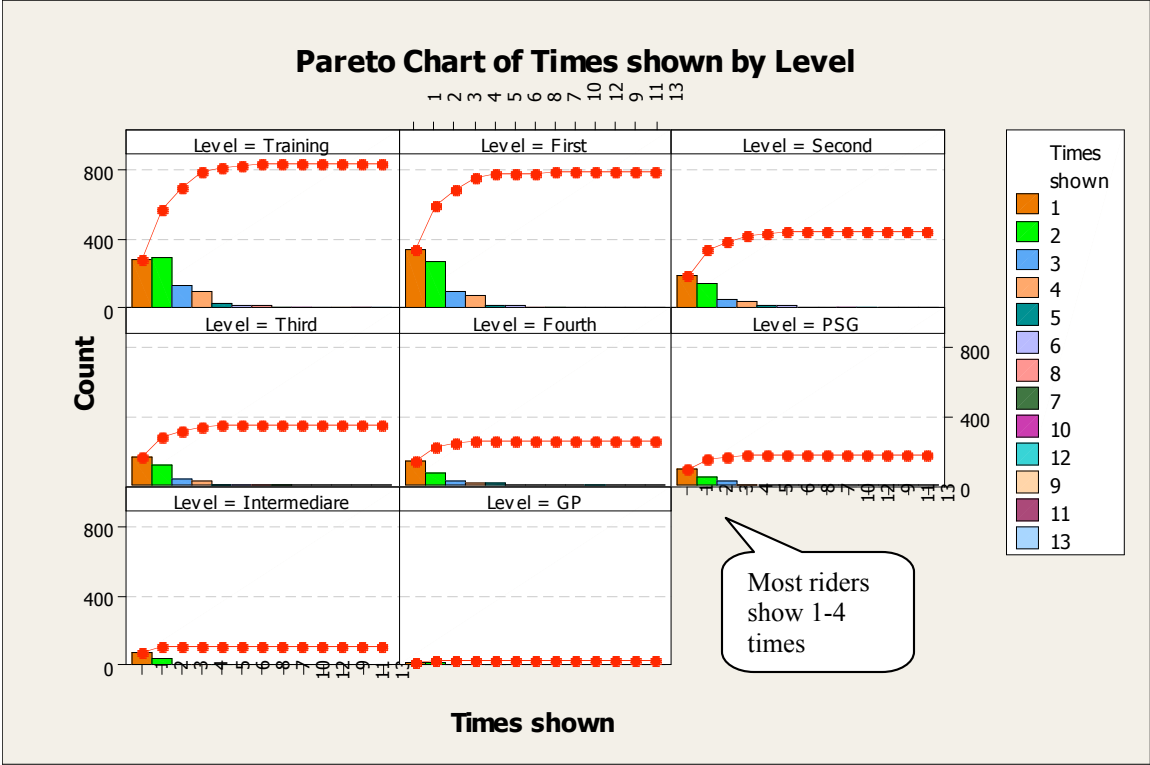


Graphical Summary of all scores

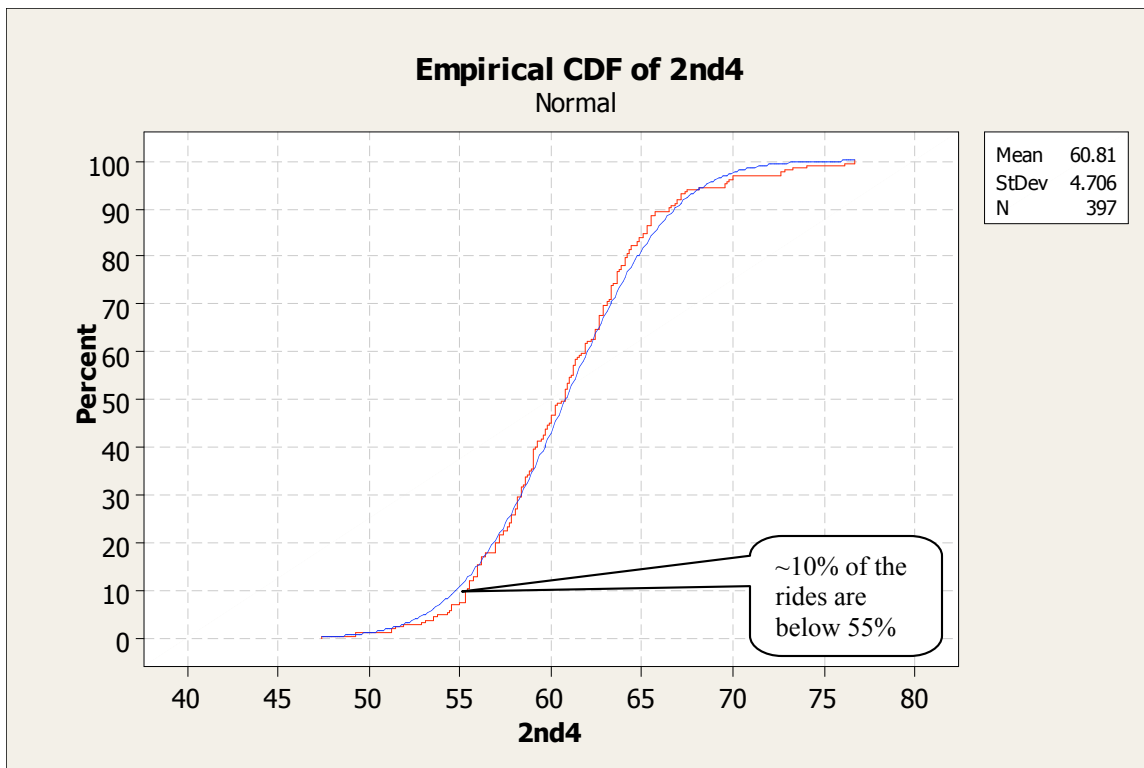
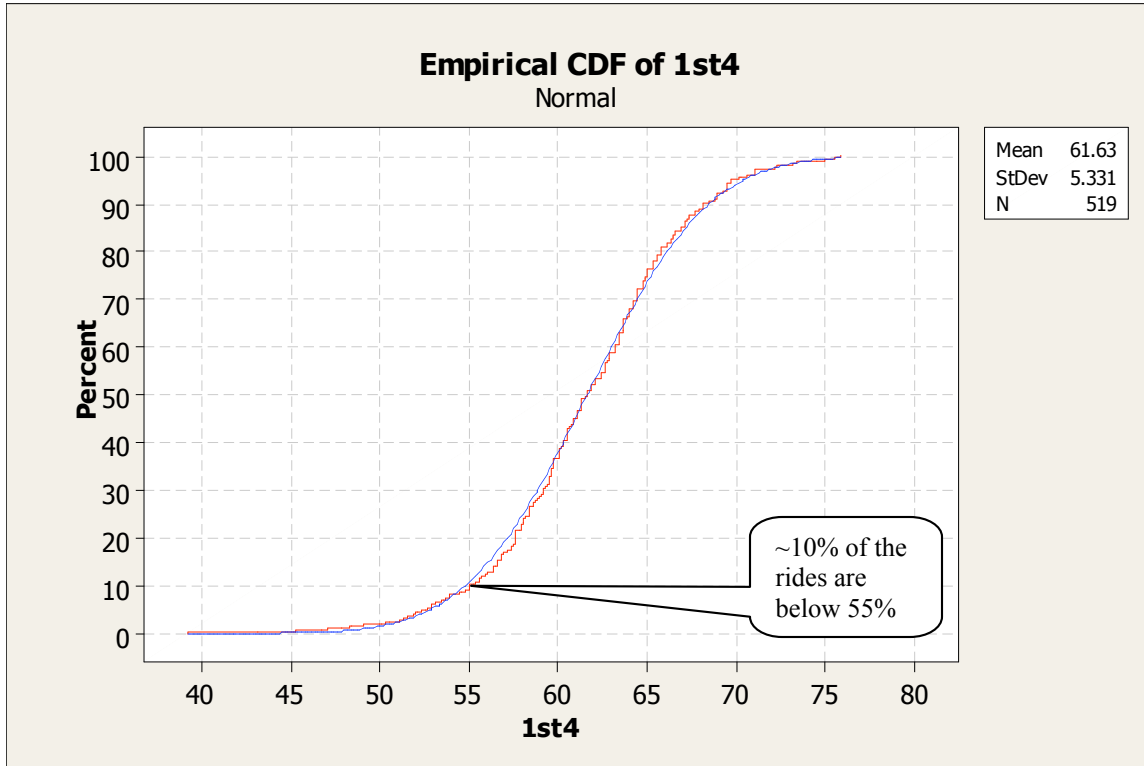


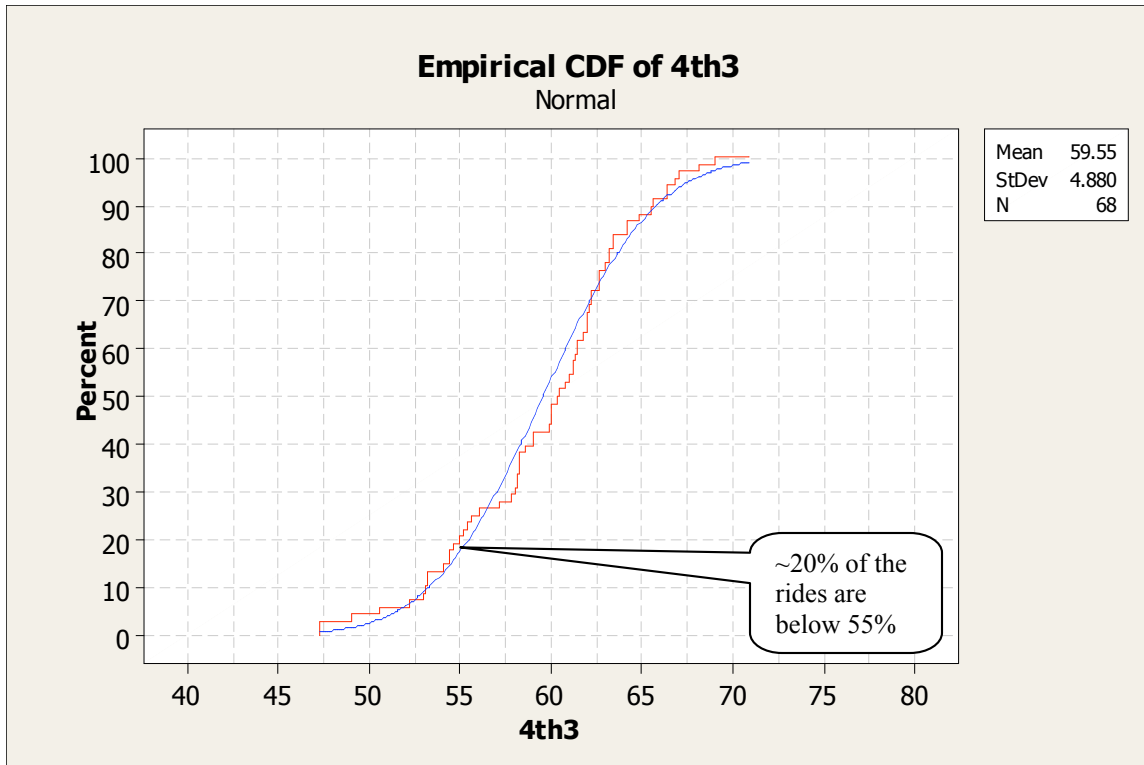
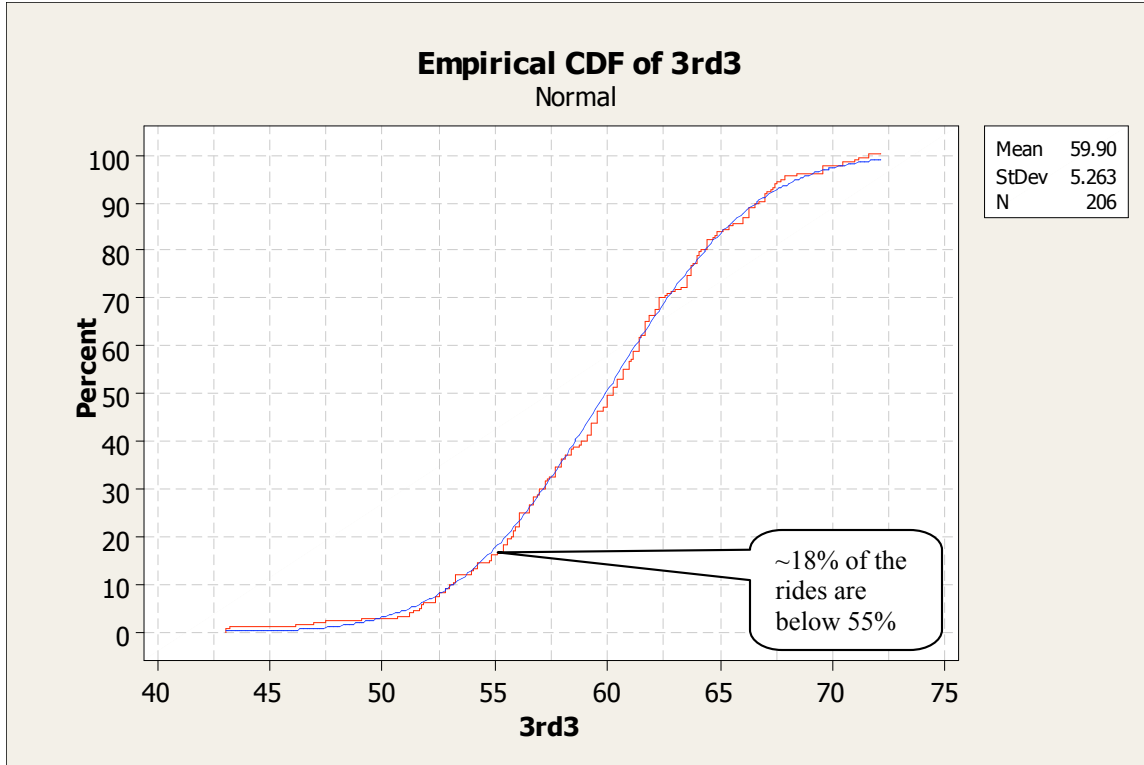


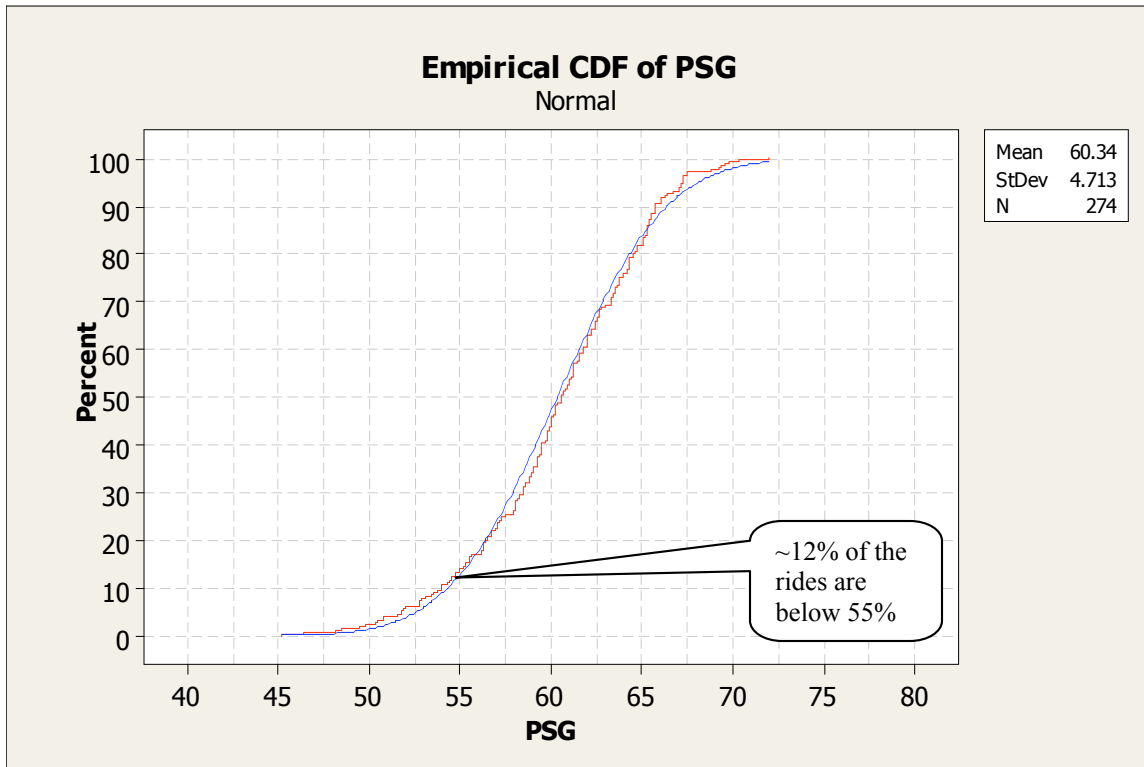




Cumulative Distribution Function (CDF): Rides by score for 1st4, 2nd4, 3rd 3, 4th 3 & PSG







About the Authors:

Dr. Rebecca Yount (dryount@yahoo.com) is a long-time PVDA, USEF, and USDF member, who has her Bronze Medal and 4th level scores toward her Silver. In her professional life, she provides psychology services to clients with developmental disabilities, including the administration and interpretation of standardized tests and the analysis of behavioral data.

Ana E. Diaz is a registered Professional Engineer with a strong background in technology and operations. She has a 25-year career at E. I. DuPont de Nemours improving quality thru statistical process control and TQM (total quality management). She finds the question of qualifying scores analogous to the quality problems she faces in her “day-job.” She is interested in applying her professional skills to help improve dressage in the US. Ms Diaz is a 4-year member of PVDA.

Mary Stydnicki Johnston is a long-time PVDA, USDF, and USEF member and previously served on the PVDA Board in the position of treasurer. She is retired from federal service, where her last assignment was as the leader of a multi-million dollar Operations Research, Modeling and Simulation office which provided decision analysis and process modeling tools used to improve government operations.

The authors have no personal stake in the outcome of the Performance Rule: Yount has her Bronze Medal and will most likely be grandfathered into any system which may eventually be approved. Diaz shows infrequently, though she was a 2nd level demo-rider for the DVCTA L-Judges March session. Johnston is re-habbing an EPM horse and not likely to show any time soon. All three authors use data and statistics professionally to enable decision making.