The CONTENTS Procedure

Alphabetic List of Variables for WORK.CDC
age exerany gender genhlth height hlthplan smoke100 weight wtdesire

The CONTENTS Procedure

Data Set Name	WORK.CDC	**Observations**	20000
Member Type	DATA	**Variables**	9
Engine	V9	**Indexes**	0
Created	11/11/2020 19:03:35	**Observation Length**	72
Last Modified	11/11/2020 19:03:35	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	23
First Data Page	1
Max Obs per Page	908
Obs in First Data Page	867
Number of Data Set Repairs	0
Filename	/tmp/SAS_work570600005A9A_10.0.2.15/SAS_work50A800005A9A_10.0.2.15/cdc.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	141429
Access Permission	rw-rw-r--
Owner Name	sasdemo
File Size	2MB
File Size (bytes)	1572864

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
8	age	Num	8	BEST12.	BEST32.
2	exerany	Num	8	BEST12.	BEST32.
9	gender	Char	1	$1.	$1.
1	genhlth	Char	9	$9.	$9.
5	height	Num	8	BEST12.	BEST32.
3	hlthplan	Num	8	BEST12.	BEST32.
4	smoke100	Num	8	BEST12.	BEST32.

The CONTENTS Procedure

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
6	weight	Num	8	BEST12.	BEST32.
7	wtdesire	Num	8	BEST12.	BEST32.

Obs	genhlth	exerany	hlthplan	smoke100	height	weight	wtdesire	age	gender
1	good	0	1	0	70	175	175	77	m
2	good	0	1	1	64	125	115	33	f
3	good	1	1	1	60	105	105	49	f
4	good	1	1	0	66	132	124	42	f
5	very good	0	1	0	61	150	130	55	f
6	very good	1	1	0	64	114	114	55	f
7	very good	1	1	0	71	194	185	31	m
8	very good	0	1	0	67	170	160	45	m
9	good	0	1	1	65	150	130	27	f
10	good	1	1	0	70	180	170	44	m

The UNIVARIATE Procedure
Variable: weight

Moments			
N	20000	**Sum Weights**	20000
Mean	169.68295	**Sum Observations**	3393659
Std Deviation	40.08097	**Variance**	1606.48415
Skewness	0.95572799	**Kurtosis**	1.99615504
Uncorrected SS	607974147	**Corrected SS**	32128076.6
Coeff Variation	23.6210945	**Std Error Mean**	0.28341526

Basic Statistical Measures			
Location		**Variability**	
Mean	169.6830	**Std Deviation**	40.08097
Median	165.0000	**Variance**	1606
Mode	160.0000	**Range**	432.00000
		Interquartile Range	50.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	598.7079	**Pr > |t|**	<.0001
Sign	**M**	10000	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.0001E8	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	500
99%	290
95%	240
90%	220
75% Q3	190
50% Median	165
25% Q1	140
10%	124
5%	115
1%	100
0% Min	68

The UNIVARIATE Procedure
Variable: weight

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
68	18743	400	2944
70	16531	400	19319
78	18065	405	15720
78	11299	495	4445
79	7614	500	1995

The FREQ Procedure

weight	Frequency	Percent	Cumulative Frequency	Cumulative Percent
68	1	0.01	1	0.01
70	1	0.01	2	0.01
78	2	0.01	4	0.02
79	1	0.01	5	0.03
80	2	0.01	7	0.04
82	2	0.01	9	0.05
83	1	0.01	10	0.05
84	3	0.02	13	0.07
85	4	0.02	17	0.09
86	1	0.01	18	0.09
88	4	0.02	22	0.11
90	12	0.06	34	0.17
92	7	0.04	41	0.21
93	6	0.03	47	0.24
94	3	0.02	50	0.25
95	22	0.11	72	0.36
96	3	0.02	75	0.38
97	6	0.03	81	0.41
98	25	0.13	106	0.53
99	5	0.03	111	0.56
100	94	0.47	205	1.03
101	6	0.03	211	1.06
102	18	0.09	229	1.15
103	30	0.15	259	1.30
104	17	0.09	276	1.38
105	112	0.56	388	1.94
106	24	0.12	412	2.06
107	33	0.17	445	2.23
108	55	0.28	500	2.50
109	14	0.07	514	2.57
110	235	1.18	749	3.75
111	7	0.04	756	3.78
112	69	0.35	825	4.13
113	34	0.17	859	4.30
114	39	0.20	898	4.49
115	244	1.22	1142	5.71

The FREQ Procedure

weight	Frequency	Percent	Cumulative Frequency	Cumulative Percent
116	33	0.17	1175	5.88
117	48	0.24	1223	6.12
118	102	0.51	1325	6.63
119	32	0.16	1357	6.79
120	440	2.20	1797	8.99
121	24	0.12	1821	9.11
122	74	0.37	1895	9.48
123	65	0.33	1960	9.80
124	58	0.29	2018	10.09
125	473	2.37	2491	12.46
126	59	0.30	2550	12.75
127	71	0.36	2621	13.11
128	125	0.63	2746	13.73
129	37	0.19	2783	13.92
130	627	3.14	3410	17.05
131	26	0.13	3436	17.18
132	110	0.55	3546	17.73
133	46	0.23	3592	17.96
134	71	0.36	3663	18.32
135	589	2.95	4252	21.26
136	60	0.30	4312	21.56
137	62	0.31	4374	21.87
138	144	0.72	4518	22.59
139	30	0.15	4548	22.74
140	794	3.97	5342	26.71
141	30	0.15	5372	26.86
142	110	0.55	5482	27.41
143	62	0.31	5544	27.72
144	41	0.21	5585	27.92
145	615	3.08	6200	31.00
146	43	0.22	6243	31.21
147	62	0.31	6305	31.52
148	111	0.56	6416	32.08
149	40	0.20	6456	32.28
150	970	4.85	7426	37.13
151	25	0.13	7451	37.26

The FREQ Procedure

weight	Frequency	Percent	Cumulative Frequency	Cumulative Percent
152	80	0.40	7531	37.66
153	47	0.24	7578	37.89
154	47	0.24	7625	38.13
155	527	2.64	8152	40.76
156	49	0.25	8201	41.00
157	48	0.24	8249	41.25
158	102	0.51	8351	41.75
159	28	0.14	8379	41.89
160	992	4.96	9371	46.85
161	19	0.10	9390	46.95
162	96	0.48	9486	47.43
163	60	0.30	9546	47.73
164	46	0.23	9592	47.96
165	692	3.46	10284	51.42
166	26	0.13	10310	51.55
167	45	0.23	10355	51.78
168	122	0.61	10477	52.39
169	28	0.14	10505	52.53
170	922	4.61	11427	57.14
171	26	0.13	11453	57.27
172	95	0.48	11548	57.74
173	50	0.25	11598	57.99
174	53	0.27	11651	58.26
175	626	3.13	12277	61.39
176	40	0.20	12317	61.59
177	22	0.11	12339	61.70
178	106	0.53	12445	62.23
179	37	0.19	12482	62.41
180	933	4.67	13415	67.08
181	11	0.06	13426	67.13
182	64	0.32	13490	67.45
183	45	0.23	13535	67.68
184	40	0.20	13575	67.88
185	577	2.89	14152	70.76
186	42	0.21	14194	70.97
187	47	0.24	14241	71.21

The FREQ Procedure

weight	Frequency	Percent	Cumulative Frequency	Cumulative Percent
188	49	0.25	14290	71.45
189	33	0.17	14323	71.62
190	715	3.58	15038	75.19
191	9	0.05	15047	75.24
192	64	0.32	15111	75.56
193	31	0.16	15142	75.71
194	31	0.16	15173	75.87
195	393	1.97	15566	77.83
196	29	0.15	15595	77.98
197	34	0.17	15629	78.15
198	70	0.35	15699	78.50
199	18	0.09	15717	78.59
200	805	4.03	16522	82.61
201	8	0.04	16530	82.65
202	24	0.12	16554	82.77
203	32	0.16	16586	82.93
204	36	0.18	16622	83.11
205	230	1.15	16852	84.26
206	19	0.10	16871	84.36
207	21	0.11	16892	84.46
208	33	0.17	16925	84.63
209	10	0.05	16935	84.68
210	431	2.16	17366	86.83
211	7	0.04	17373	86.87
212	45	0.23	17418	87.09
213	9	0.05	17427	87.14
214	14	0.07	17441	87.21
215	206	1.03	17647	88.24
216	18	0.09	17665	88.33
217	11	0.06	17676	88.38
218	34	0.17	17710	88.55
219	11	0.06	17721	88.61
220	376	1.88	18097	90.49
221	1	0.01	18098	90.49
222	18	0.09	18116	90.58
223	7	0.04	18123	90.62

The FREQ Procedure

weight	Frequency	Percent	Cumulative Frequency	Cumulative Percent
224	8	0.04	18131	90.66
225	196	0.98	18327	91.64
226	8	0.04	18335	91.68
227	10	0.05	18345	91.73
228	22	0.11	18367	91.84
229	1	0.01	18368	91.84
230	268	1.34	18636	93.18
231	2	0.01	18638	93.19
232	9	0.05	18647	93.24
233	4	0.02	18651	93.26
234	8	0.04	18659	93.30
235	137	0.69	18796	93.98
236	11	0.06	18807	94.04
237	5	0.03	18812	94.06
238	12	0.06	18824	94.12
239	4	0.02	18828	94.14
240	204	1.02	19032	95.16
241	4	0.02	19036	95.18
242	10	0.05	19046	95.23
243	4	0.02	19050	95.25
244	1	0.01	19051	95.26
245	69	0.35	19120	95.60
246	6	0.03	19126	95.63
247	2	0.01	19128	95.64
248	18	0.09	19146	95.73
249	2	0.01	19148	95.74
250	202	1.01	19350	96.75
252	9	0.05	19359	96.80
253	4	0.02	19363	96.82
254	2	0.01	19365	96.83
255	27	0.14	19392	96.96
256	4	0.02	19396	96.98
257	8	0.04	19404	97.02
258	5	0.03	19409	97.05
260	104	0.52	19513	97.57
262	6	0.03	19519	97.60

The FREQ Procedure

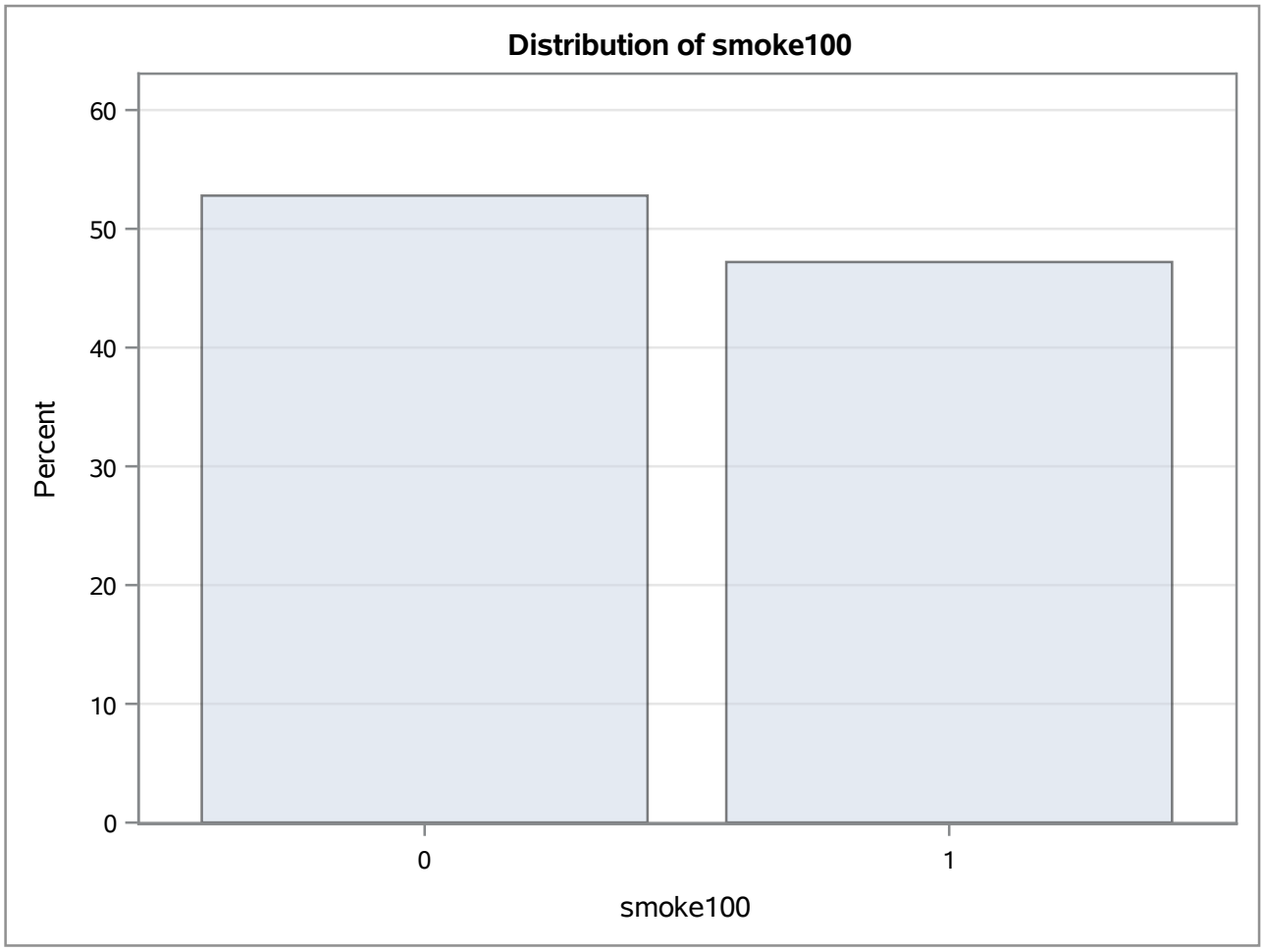
weight	Frequency	Percent	Cumulative Frequency	Cumulative Percent
263	4	0.02	19523	97.62
265	36	0.18	19559	97.80
267	4	0.02	19563	97.82
268	4	0.02	19567	97.84
270	56	0.28	19623	98.12
271	1	0.01	19624	98.12
272	3	0.02	19627	98.14
273	1	0.01	19628	98.14
274	3	0.02	19631	98.16
275	40	0.20	19671	98.36
276	2	0.01	19673	98.37
278	6	0.03	19679	98.40
279	2	0.01	19681	98.41
280	57	0.29	19738	98.69
282	2	0.01	19740	98.70
283	2	0.01	19742	98.71
285	21	0.11	19763	98.82
286	3	0.02	19766	98.83
287	2	0.01	19768	98.84
290	40	0.20	19808	99.04
292	2	0.01	19810	99.05
294	1	0.01	19811	99.06
295	11	0.06	19822	99.11
296	1	0.01	19823	99.12
297	1	0.01	19824	99.12
298	3	0.02	19827	99.14
300	70	0.35	19897	99.49
305	4	0.02	19901	99.51
308	1	0.01	19902	99.51
309	1	0.01	19903	99.52
310	12	0.06	19915	99.58
313	1	0.01	19916	99.58
315	7	0.04	19923	99.62
318	1	0.01	19924	99.62
319	1	0.01	19925	99.63
320	12	0.06	19937	99.69

The FREQ Procedure

weight	Frequency	Percent	Cumulative Frequency	Cumulative Percent
324	1	0.01	19938	99.69
325	3	0.02	19941	99.71
327	1	0.01	19942	99.71
328	1	0.01	19943	99.72
330	6	0.03	19949	99.75
340	6	0.03	19955	99.78
344	1	0.01	19956	99.78
348	1	0.01	19957	99.79
350	21	0.11	19978	99.89
360	2	0.01	19980	99.90
362	2	0.01	19982	99.91
364	1	0.01	19983	99.92
370	1	0.01	19984	99.92
371	1	0.01	19985	99.93
380	4	0.02	19989	99.95
385	2	0.01	19991	99.96
390	1	0.01	19992	99.96
400	5	0.03	19997	99.99
405	1	0.01	19998	99.99
495	1	0.01	19999	100.00
500	1	0.01	20000	100.00

The FREQ Procedure

smoke100	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	10559	52.80	10559	52.80
1	9441	47.21	20000	100.00



The UNIVARIATE Procedure
Variable: height

Moments			
N	20000	**Sum Weights**	20000
Mean	67.1829	**Sum Observations**	1343658
Std Deviation	4.12595429	**Variance**	17.0234988
Skewness	0.1036124	**Kurtosis**	-0.3783851
Uncorrected SS	90611294	**Corrected SS**	340452.952
Coeff Variation	6.14137569	**Std Error Mean**	0.0291749

Basic Statistical Measures			
Location		**Variability**	
Mean	67.18290	**Std Deviation**	4.12595
Median	67.00000	**Variance**	17.02350
Mode	66.00000	**Range**	45.00000
		Interquartile Range	6.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	2302.763	**Pr > |t|**	<.0001
Sign	**M**	10000	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.0001E8	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	93
99%	76
95%	74
90%	73
75% Q3	70
50% Median	67
25% Q1	64
10%	62
5%	61
1%	59
0% Min	48

The UNIVARIATE Procedure
Variable: height

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
48	15465	82	10160
48	5412	82	10322
49	8871	83	3691
50	3905	84	18817
51	11948	93	17534

The UNIVARIATE Procedure
Variable: age

Moments			
N	20000	**Sum Weights**	20000
Mean	45.06825	**Sum Observations**	901365
Std Deviation	17.1926895	**Variance**	295.588571
Skewness	0.45170032	**Kurtosis**	-0.6548259
Uncorrected SS	46534419	**Corrected SS**	5911475.84
Coeff Variation	38.1481186	**Std Error Mean**	0.12157067

Basic Statistical Measures			
Location		**Variability**	
Mean	45.06825	**Std Deviation**	17.19269
Median	43.00000	**Variance**	295.58857
Mode	40.00000	**Range**	81.00000
		Interquartile Range	26.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	370.7165	**Pr > |t|**	<.0001
Sign	**M**	10000	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.0001E8	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	99
99%	84
95%	76
90%	71
75% Q3	57
50% Median	43
25% Q1	31
10%	24
5%	21
1%	18
0% Min	18

The UNIVARIATE Procedure
Variable: age

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
18	19885	95	16084
18	19860	96	17051
18	19832	97	10350
18	19706	99	900
18	19622	99	6710

The FREQ Procedure

gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
f	10431	52.16	10431	52.16
m	9569	47.85	20000	100.00

The FREQ Procedure

exerany	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	5086	25.43	5086	25.43
1	14914	74.57	20000	100.00

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of gender by smoke100

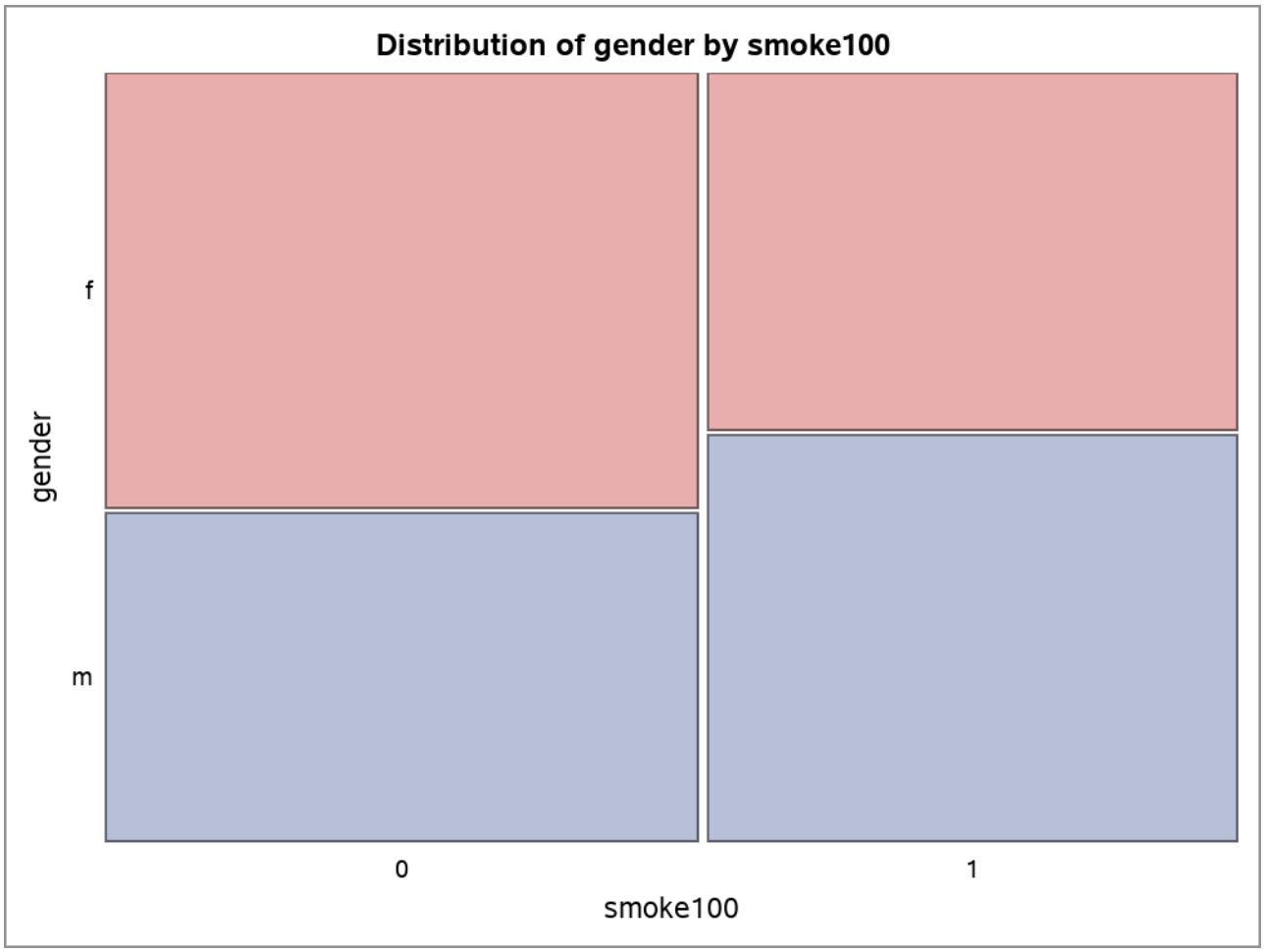
gender	smoke100 0	smoke100 1	Total
f	6012 30.06 57.64 56.94	4419 22.10 42.36 46.81	10431 52.16
m	4547 22.74 47.52 43.06	5022 25.11 52.48 53.19	9569 47.85
Total	10559 52.80	9441 47.21	20000 100.00

The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of gender by smoke100

gender	smoke100 0	smoke100 1	Total
f	6012 30.06 57.64 56.94	4419 22.10 42.36 46.81	10431 52.16
m	4547 22.74 47.52 43.06	5022 25.11 52.48 53.19	9569 47.85
Total	10559 52.80	9441 47.21	20000 100.00



Answer
The chunk of the females that smoke less than 100 cigarettes is larger than the piece of males. Also, the chunk of males that smoke more than 100 cigarettes is larger than the piece of females.

Obs	genhlth	exerany	hlthplan	smoke100	height	weight	wtdesire	age	gender
1	good	0	1	0	70	175	175	77	m
2	good	0	1	1	64	125	115	33	f
3	good	1	1	1	60	105	105	49	f
4	good	1	1	0	66	132	124	42	f
5	very good	0	1	0	61	150	130	55	f
6	very good	1	1	0	64	114	114	55	f
7	very good	1	1	0	71	194	185	31	m
8	very good	0	1	0	67	170	160	45	m
9	good	0	1	1	65	150	130	27	f
10	good	1	1	0	70	180	170	44	m

The CONTENTS Procedure

Data Set Name	WORK.CDC	**Observations**	20000
Member Type	DATA	**Variables**	9
Engine	V9	**Indexes**	0
Created	11/11/2020 19:03:35	**Observation Length**	72
Last Modified	11/11/2020 19:03:35	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	23
First Data Page	1
Max Obs per Page	908
Obs in First Data Page	867
Number of Data Set Repairs	0
Filename	/tmp/SAS_work570600005A9A_10.0.2.15/SAS_work50A800005A9A_10.0.2.15/cdc.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	141429
Access Permission	rw-rw-r--
Owner Name	sasdemo
File Size	2MB
File Size (bytes)	1572864

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
8	age	Num	8	BEST12.	BEST32.
2	exerany	Num	8	BEST12.	BEST32.
9	gender	Char	1	$1.	$1.
1	genhlth	Char	9	$9.	$9.
5	height	Num	8	BEST12.	BEST32.
3	hlthplan	Num	8	BEST12.	BEST32.
4	smoke100	Num	8	BEST12.	BEST32.

The CONTENTS Procedure

Alphabetic List of Variables and Attributes					
#	**Variable**	**Type**	**Len**	**Format**	**Informat**
6	weight	Num	8	BEST12.	BEST32.
7	wtdesire	Num	8	BEST12.	BEST32.

Obs	genhlth	exerany	hlthplan	smoke100	height	weight	wtdesire	age	gender
1	good	0	1	0	70	175	175	77	m
2	very good	1	1	0	71	194	185	31	m
3	very good	0	1	0	67	170	160	45	m
4	good	1	1	0	70	180	170	44	m
5	excellent	1	1	1	69	186	175	46	m
6	fair	1	1	1	69	168	148	62	m
7	excellent	1	1	1	70	170	170	69	m
8	good	1	1	1	73	185	175	79	m
9	good	0	0	1	67	156	150	47	m
10	fair	0	1	1	71	185	185	76	m

Obs	genhlth	exerany	hlthplan	smoke100	height	weight	wtdesire	age	gender	wtkilos	wtdiff	now
1	Good	0	1	0	70	175	175	77	m	79.3786	0	22230
2	Good	0	1	1	64	125	115	33	f	56.6990	10	22230
3	Good	1	1	1	60	105	105	49	f	47.6272	0	22230
4	Good	1	1	0	66	132	124	42	f	59.8741	8	22230
5	Very Good	0	1	0	61	150	130	55	f	68.0388	20	22230
6	Very Good	1	1	0	64	114	114	55	f	51.7095	0	22230
7	Very Good	1	1	0	71	194	185	31	m	87.9968	9	22230
8	Very Good	0	1	0	67	170	160	45	m	77.1106	10	22230
9	Good	0	1	1	65	150	130	27	f	68.0388	20	22230
10	Good	1	1	0	70	180	170	44	m	81.6466	10	22230

Obs	genhlth	exerany	hlthplan	smoke100	height	weight	wtdesire	age	gender
1	good	0	1	0	70	175	175	77	m
2	very good	1	1	0	71	194	185	31	m
3	very good	0	1	0	67	170	160	45	m
4	good	1	1	0	70	180	170	44	m
5	excellent	1	1	1	69	186	175	46	m
6	fair	1	1	1	69	168	148	62	m
7	excellent	1	0	1	66	185	220	21	m
8	excellent	1	1	1	70	170	170	69	m
9	fair	1	0	0	69	170	170	23	m
10	good	1	1	1	73	185	175	79	m

Obs	genhlth	exerany	hlthplan	smoke100	height	weight	wtdesire	age	gender
1	good	0	1	0	70	175	175	77	m
2	good	0	1	1	64	125	115	33	f
3	good	1	1	1	60	105	105	49	f
4	good	1	1	0	66	132	124	42	f
5	very good	0	1	0	61	150	130	55	f
6	very good	1	1	0	64	114	114	55	f
7	very good	1	1	0	71	194	185	31	m
8	very good	0	1	0	67	170	160	45	m
9	good	1	1	0	70	180	170	44	m
10	excellent	1	1	1	69	186	175	46	m

Obs	genhlth	exerany	hlthplan	smoke100	height	weight	wtdesire	age	gender
1	good	0	1	0	70	175	175	77	m
2	good	0	1	1	64	125	115	33	f
3	good	1	1	1	60	105	105	49	f
4	good	1	1	0	66	132	124	42	f
5	very good	0	1	0	61	150	130	55	f
6	very good	1	1	0	64	114	114	55	f
7	very good	1	1	0	71	194	185	31	m
8	very good	0	1	0	67	170	160	45	m
9	good	1	1	0	70	180	170	44	m
10	excellent	1	1	1	69	186	175	46	m

Obs	genhlth	exerany	hlthplan	smoke100	height	weight	wtdesire	age	gender
1	excellent	1	0	1	66	185	220	21	m
2	very good	1	0	1	70	160	140	18	f
3	excellent	1	1	1	74	175	200	22	m
4	good	1	1	1	64	190	140	20	f
5	very good	1	1	1	62	92	92	21	f
6	very good	1	0	1	64	125	115	22	f
7	fair	0	1	1	71	185	185	20	m
8	fair	1	1	1	72	185	170	19	m
9	excellent	1	0	1	63	105	100	19	m
10	fair	1	1	1	71	185	150	18	m

The UNIVARIATE Procedure
Variable: weight

Moments			
N	20000	**Sum Weights**	20000
Mean	169.68295	**Sum Observations**	3393659
Std Deviation	40.08097	**Variance**	1606.48415
Skewness	0.95572799	**Kurtosis**	1.99615504
Uncorrected SS	607974147	**Corrected SS**	32128076.6
Coeff Variation	23.6210945	**Std Error Mean**	0.28341526

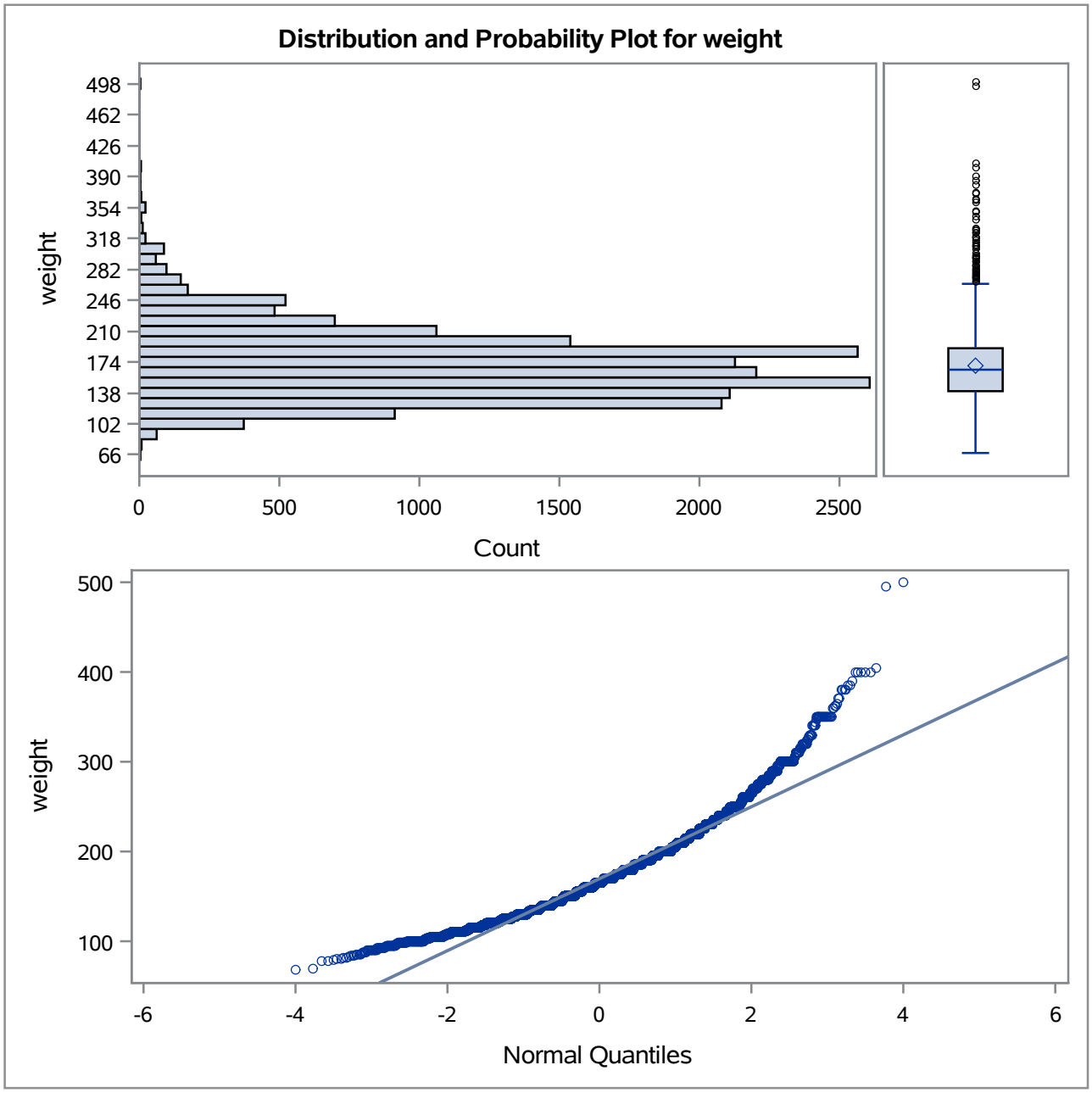
Basic Statistical Measures			
Location		**Variability**	
Mean	169.6830	**Std Deviation**	40.08097
Median	165.0000	**Variance**	1606
Mode	160.0000	**Range**	432.00000
		Interquartile Range	50.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	598.7079	**Pr > |t|**	<.0001
Sign	**M**	10000	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.0001E8	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	500
99%	290
95%	240
90%	220
75% Q3	190
50% Median	165
25% Q1	140
10%	124
5%	115
1%	100
0% Min	68

The UNIVARIATE Procedure
Variable: weight

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
68	18743	400	2944
70	16531	400	19319
78	18065	405	15720
78	11299	495	4445
79	7614	500	1995



The UNIVARIATE Procedure
Variable: weight
gender = f

Moments			
N	10431	**Sum Weights**	10431
Mean	151.666187	**Sum Observations**	1582030
Std Deviation	34.2975191	**Variance**	1176.31981
Skewness	1.38530136	**Kurtosis**	3.77381659
Uncorrected SS	252209474	**Corrected SS**	12269015.7
Coeff Variation	22.6138203	**Std Error Mean**	0.33581472

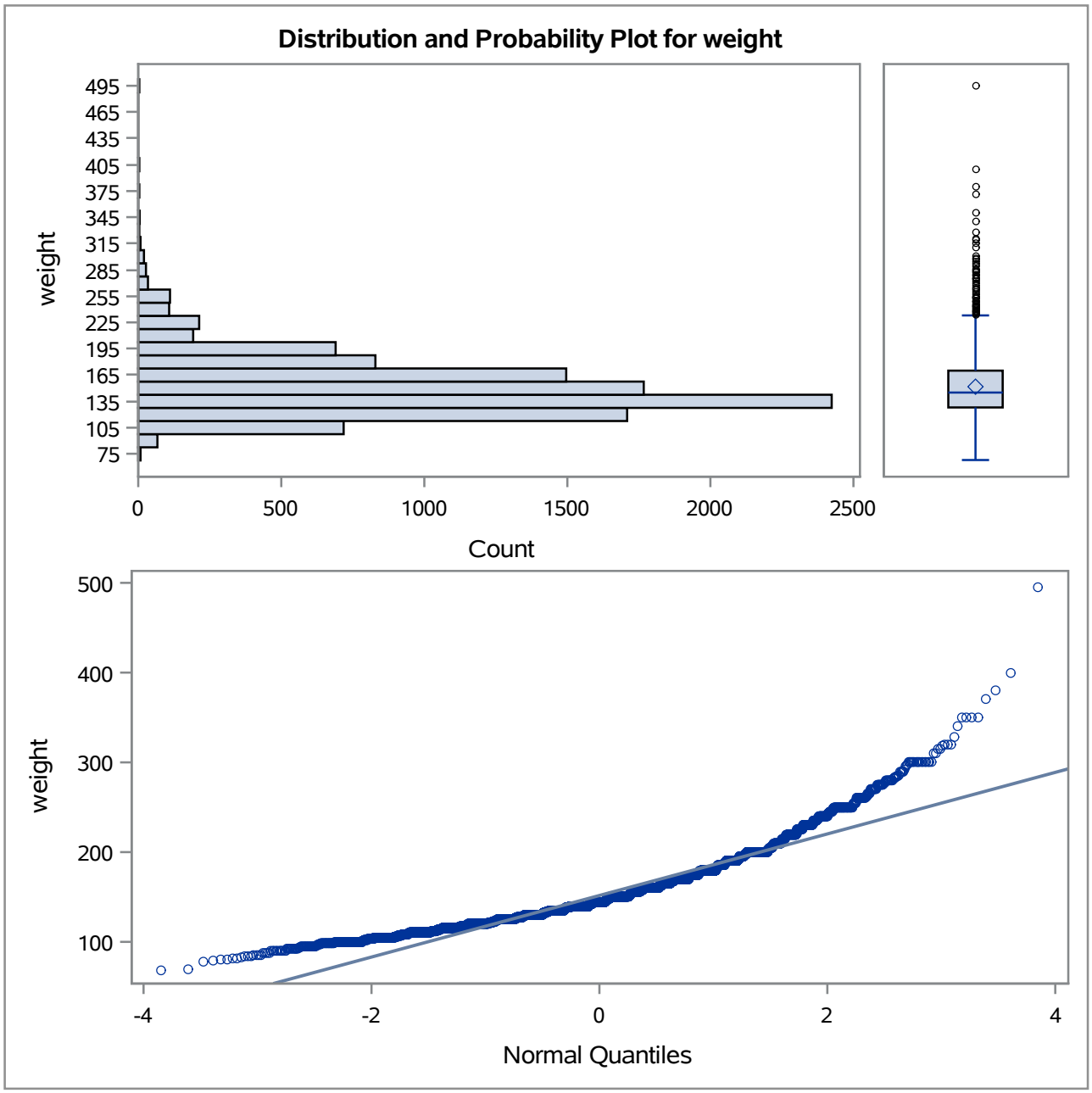
Basic Statistical Measures			
Location		**Variability**	
Mean	151.6662	**Std Deviation**	34.29752
Median	145.0000	**Variance**	1176
Mode	140.0000	**Range**	427.00000
		Interquartile Range	42.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	451.6365	**Pr > |t|**	<.0001
Sign	**M**	5215.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	27204048	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	495
99%	260
95%	220
90%	198
75% Q3	170
50% Median	145
25% Q1	128
10%	115
5%	110
1%	100
0% Min	68

The UNIVARIATE Procedure
Variable: weight
gender = f

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
68	18743	350	13607
70	16531	371	4612
78	11299	380	7160
79	7614	400	19319
80	15673	495	4445



The UNIVARIATE Procedure
Variable: weight
gender = m

Moments			
N	9569	**Sum Weights**	9569
Mean	189.322709	**Sum Observations**	1811629
Std Deviation	36.5503551	**Variance**	1335.92846
Skewness	1.16960127	**Kurtosis**	3.04520553
Uncorrected SS	355764673	**Corrected SS**	12782163.5
Coeff Variation	19.3058484	**Std Error Mean**	0.37364427

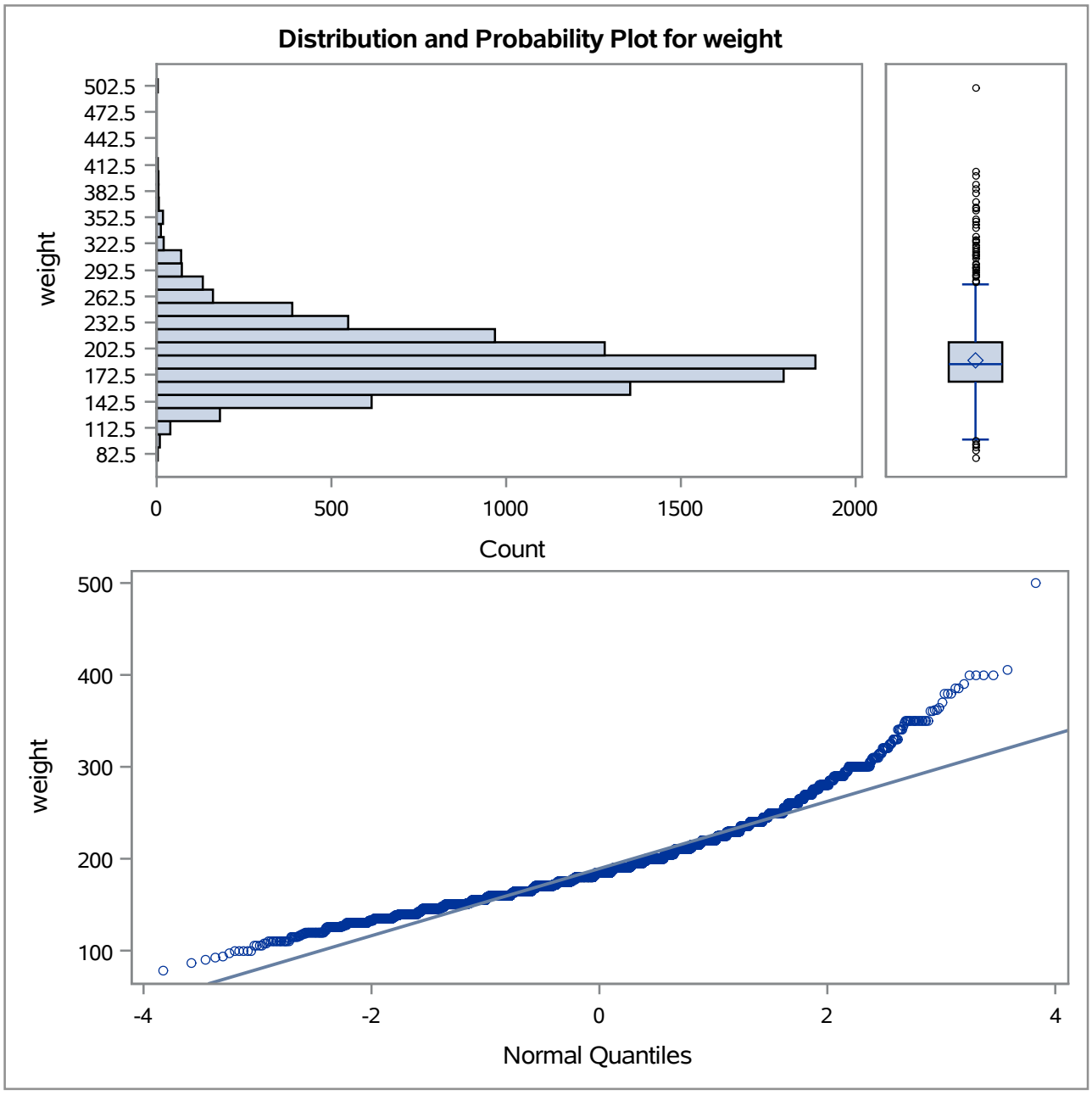
Basic Statistical Measures			
Location		**Variability**	
Mean	189.3227	**Std Deviation**	36.55036
Median	185.0000	**Variance**	1336
Mode	180.0000	**Range**	422.00000
		Interquartile Range	45.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	506.6924	**Pr > |t|**	<.0001
Sign	**M**	4784.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	22893833	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	500
99%	300
95%	256
90%	235
75% Q3	210
50% Median	185
25% Q1	165
10%	150
5%	140
1%	125
0% Min	78

The UNIVARIATE Procedure
Variable: weight
gender = m

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
78	18065	400	1279
86	15967	400	2659
90	303	400	2944
93	9558	405	15720
94	11333	500	1995



The UNIVARIATE Procedure
Variable: weight

gender=f

Moments			
N	10431	**Sum Weights**	10431
Mean	151.666187	**Sum Observations**	1582030
Std Deviation	34.2975191	**Variance**	1176.31981
Skewness	1.38530136	**Kurtosis**	3.77381659
Uncorrected SS	252209474	**Corrected SS**	12269015.7
Coeff Variation	22.6138203	**Std Error Mean**	0.33581472

Basic Statistical Measures			
Location		**Variability**	
Mean	151.6662	**Std Deviation**	34.29752
Median	145.0000	**Variance**	1176
Mode	140.0000	**Range**	427.00000
		Interquartile Range	42.00000

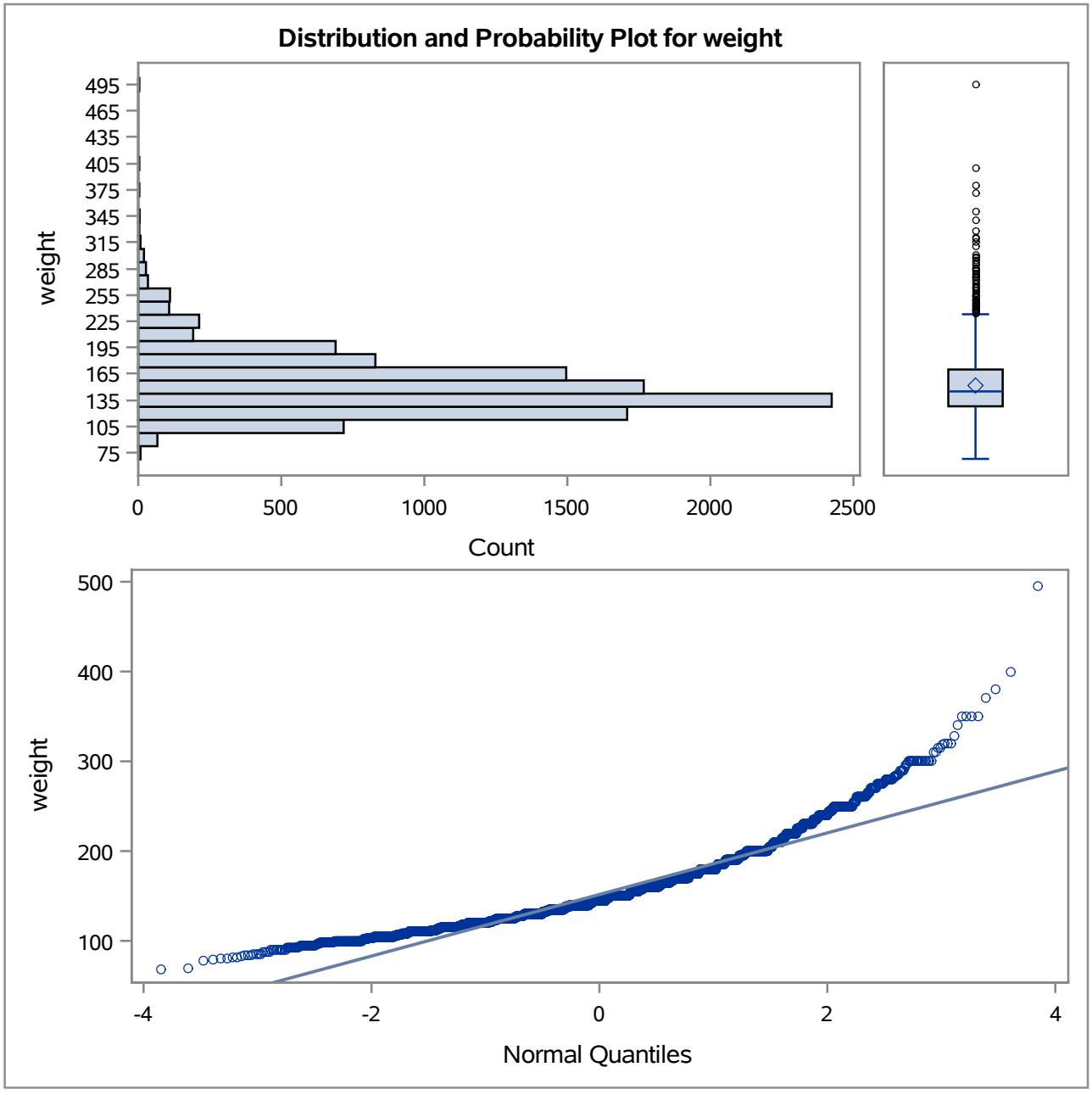
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	451.6365	**Pr > |t|**	<.0001
Sign	**M**	5215.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	27204048	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	495
99%	260
95%	220
90%	198
75% Q3	170
50% Median	145
25% Q1	128
10%	115
5%	110
1%	100
0% Min	68

The UNIVARIATE Procedure
Variable: weight

gender=f

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
68	9739	350	7051
70	8592	371	2354
78	5860	380	3663
79	3905	400	10060
80	8154	495	2284



The UNIVARIATE Procedure
Variable: weight

gender=m

Moments			
N	9569	**Sum Weights**	9569
Mean	189.322709	**Sum Observations**	1811629
Std Deviation	36.5503551	**Variance**	1335.92846
Skewness	1.16960127	**Kurtosis**	3.04520553
Uncorrected SS	355764673	**Corrected SS**	12782163.5
Coeff Variation	19.3058484	**Std Error Mean**	0.37364427

Basic Statistical Measures			
Location		**Variability**	
Mean	189.3227	**Std Deviation**	36.55036
Median	185.0000	**Variance**	1336
Mode	180.0000	**Range**	422.00000
		Interquartile Range	45.00000

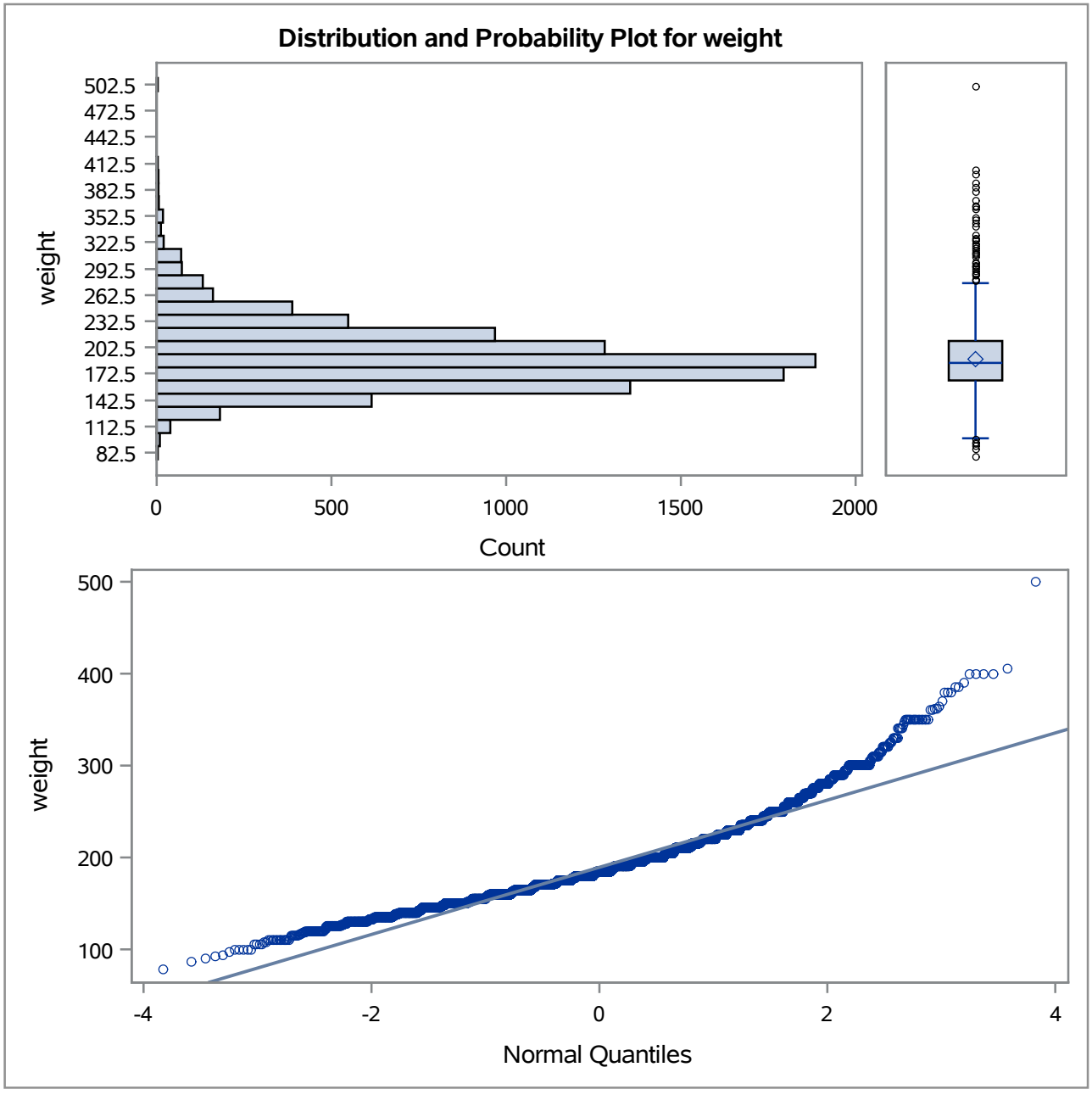
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	506.6924	**Pr > |t|**	<.0001
Sign	**M**	4784.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	22893833	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	500
99%	300
95%	256
90%	235
75% Q3	210
50% Median	185
25% Q1	165
10%	150
5%	140
1%	125
0% Min	78

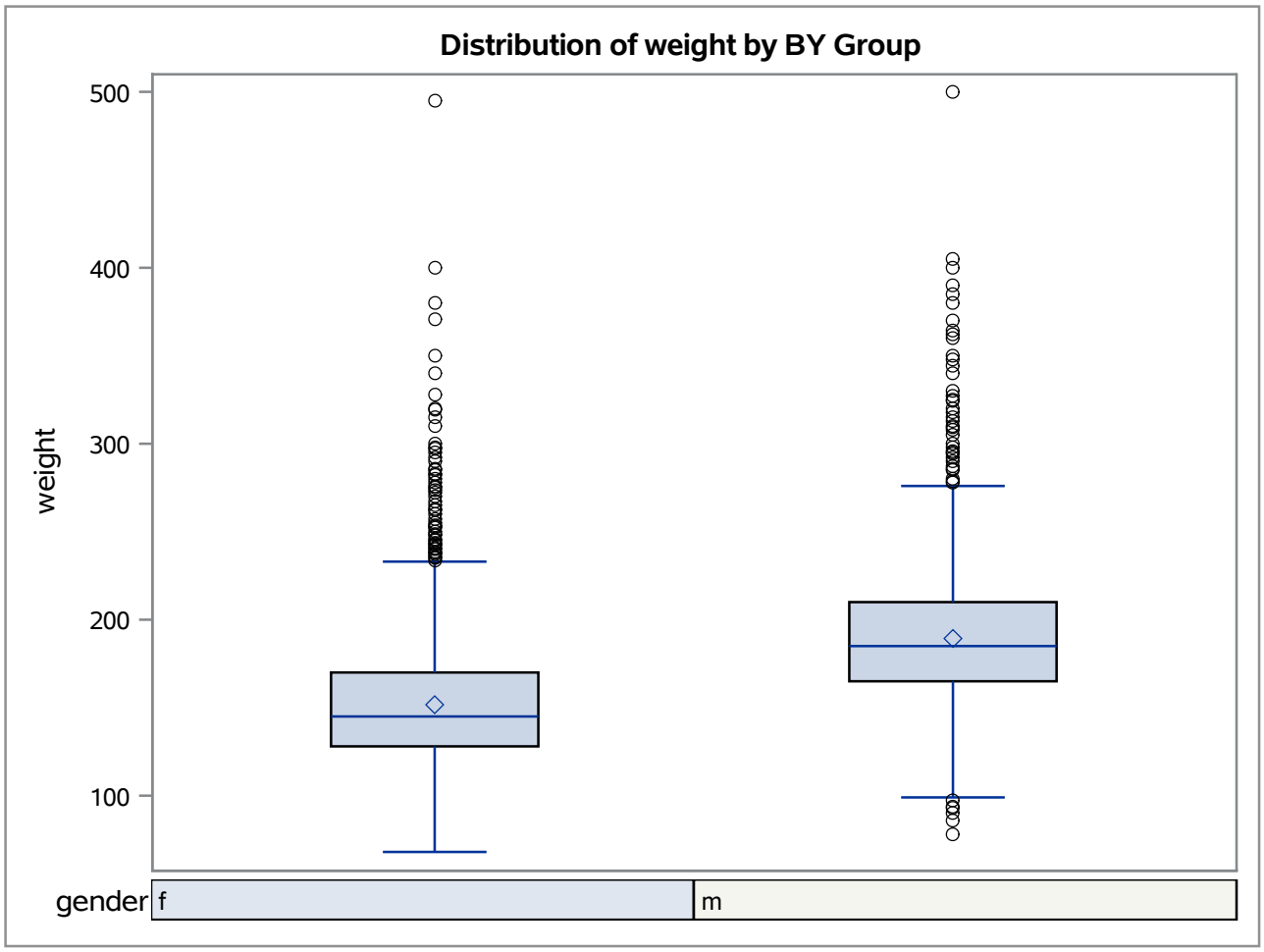
The UNIVARIATE Procedure
Variable: weight

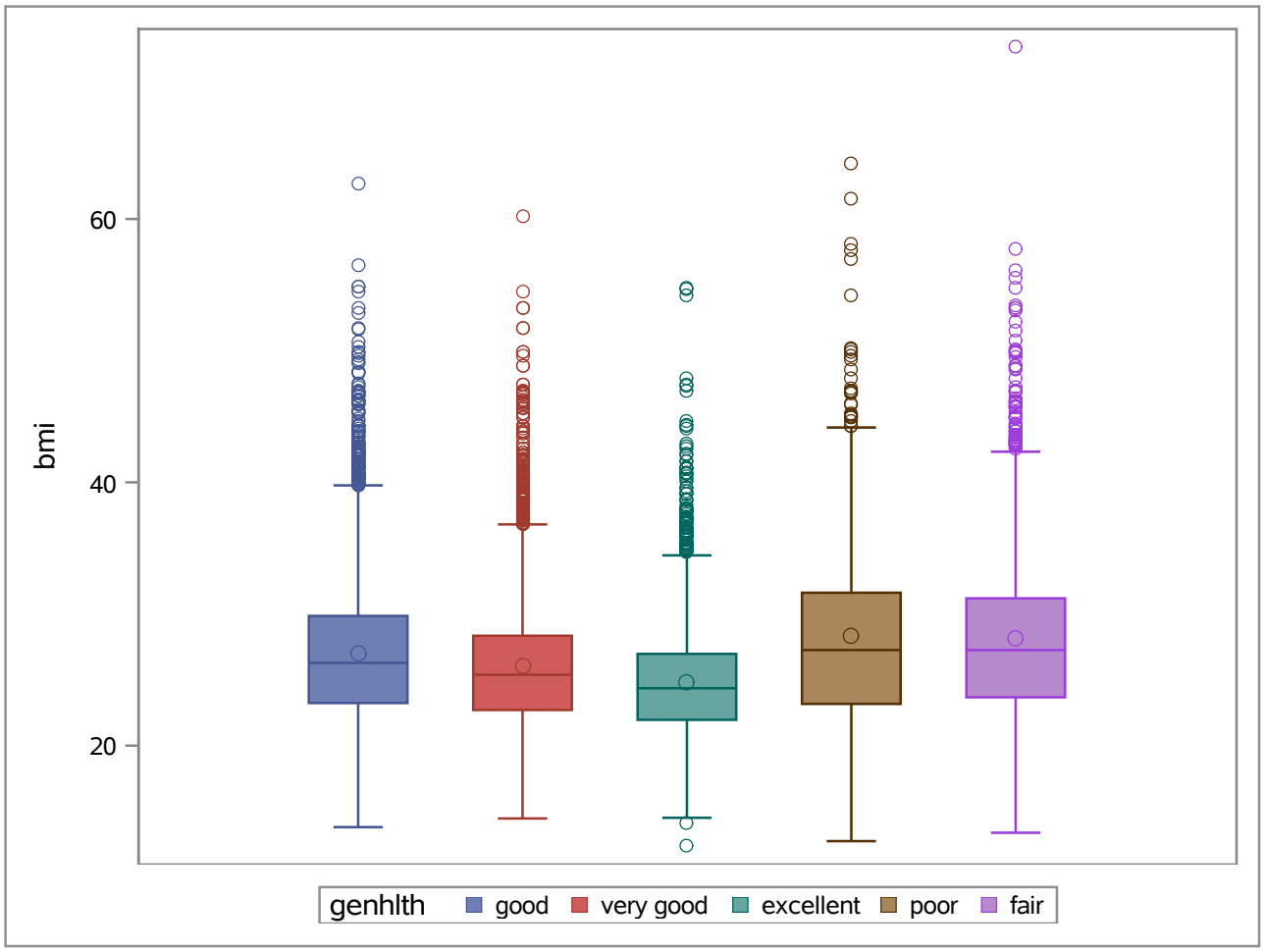
gender=m

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
78	19127	400	11036
86	18099	400	11725
90	10578	400	11867
93	15074	405	17968
94	15887	500	11392

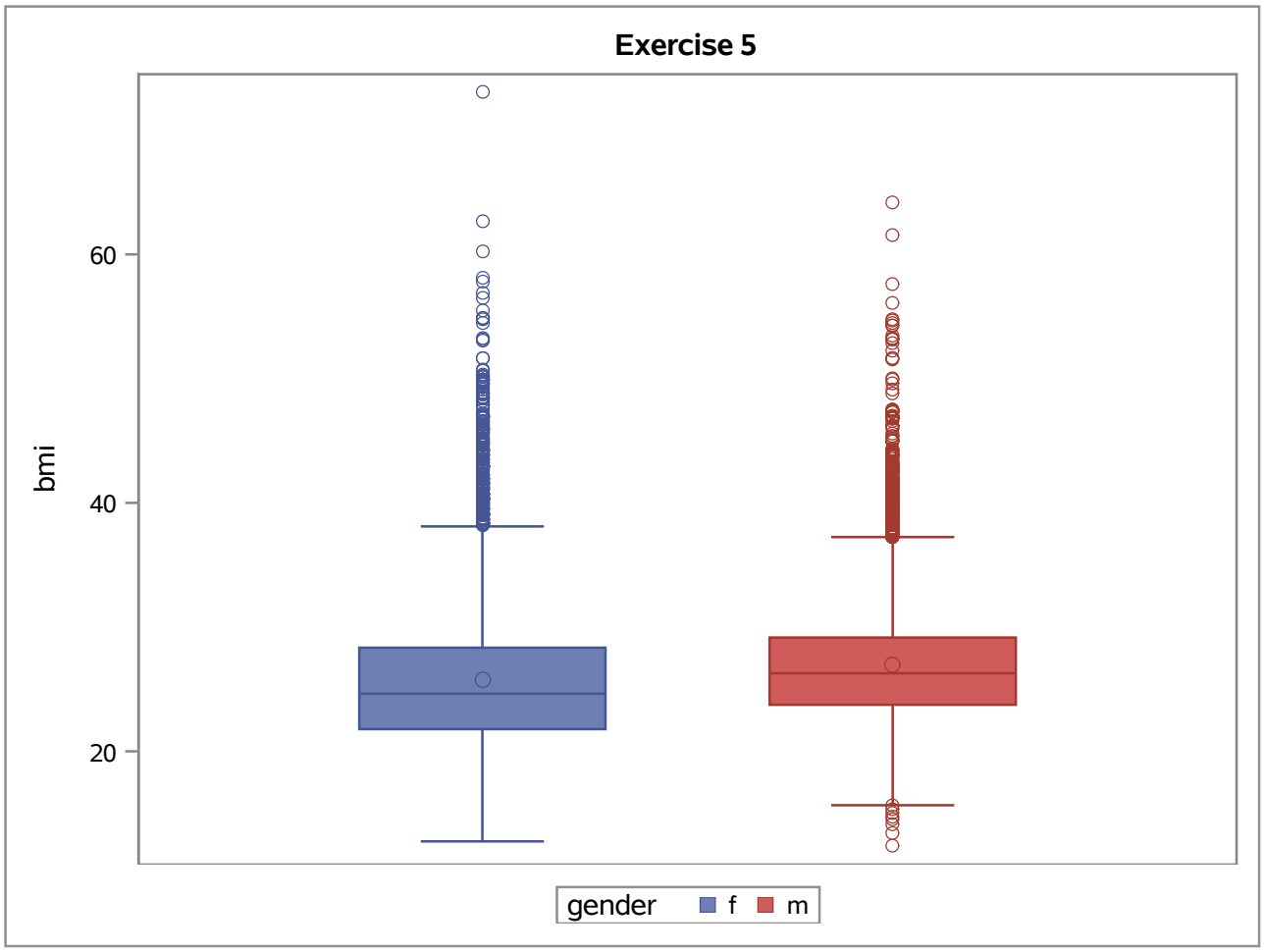


The UNIVARIATE Procedure




bmi
60
40
20
genhlth
good
very good
excellent
poor
fair

Answer
The box plot shows how the bmi varies at every value of the genhlth variable.


Exercise 5
bmi
60
40
20
gender
f
m

Answer
Gender is a reliable choice for BMI. By definition, the BMI is related to weight and height, values that are influenced by the fact that a person is a man or a woman.

The UNIVARIATE Procedure
Variable: bmi

Moments			
N	20000	**Sum Weights**	20000
Mean	26.3069252	**Sum Observations**	526138.504
Std Deviation	5.21810488	**Variance**	27.2286186
Skewness	1.27589014	**Kurtosis**	3.32068097
Uncorrected SS	14385631.4	**Corrected SS**	544545.143
Coeff Variation	19.83548	**Std Error Mean**	0.03689757

Basic Statistical Measures			
Location		**Variability**	
Mean	26.30693	**Std Deviation**	5.21810
Median	25.60354	**Variance**	27.22862
Mode	27.12191	**Range**	60.69029
		Interquartile Range	6.17852

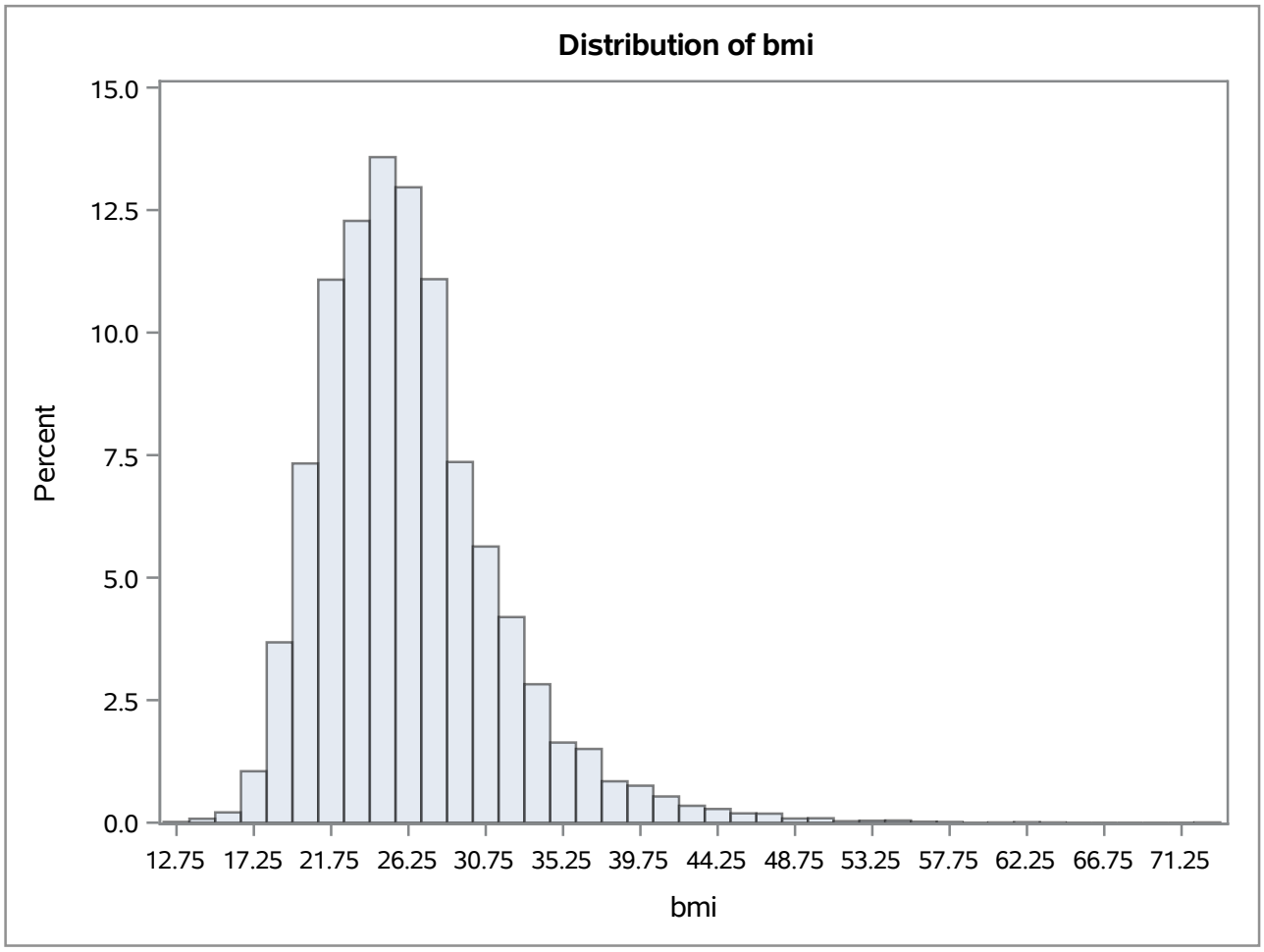
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	712.9717	**Pr > |t|**	<.0001
Sign	**M**	10000	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.0001E8	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	73.0907
99%	43.5553
95%	35.9509
90%	32.8802
75% Q3	28.8862
50% Median	25.6035
25% Q1	22.7077
10%	20.5957
5%	19.4835
1%	17.7123
0% Min	12.4005

The UNIVARIATE Procedure
Variable: bmi

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
12.4005	15887	60.2217	443
12.7495	3905	61.5733	17968
13.3872	19127	62.6420	10060
13.5565	8698	64.1892	11392
13.8104	2493	73.0907	2284

The UNIVARIATE Procedure



The UNIVARIATE Procedure
Variable: bmi

Moments			
N	20000	**Sum Weights**	20000
Mean	26.3069252	**Sum Observations**	526138.504
Std Deviation	5.21810488	**Variance**	27.2286186
Skewness	1.27589014	**Kurtosis**	3.32068097
Uncorrected SS	14385631.4	**Corrected SS**	544545.143
Coeff Variation	19.83548	**Std Error Mean**	0.03689757

Basic Statistical Measures			
Location		**Variability**	
Mean	26.30693	**Std Deviation**	5.21810
Median	25.60354	**Variance**	27.22862
Mode	27.12191	**Range**	60.69029
		Interquartile Range	6.17852

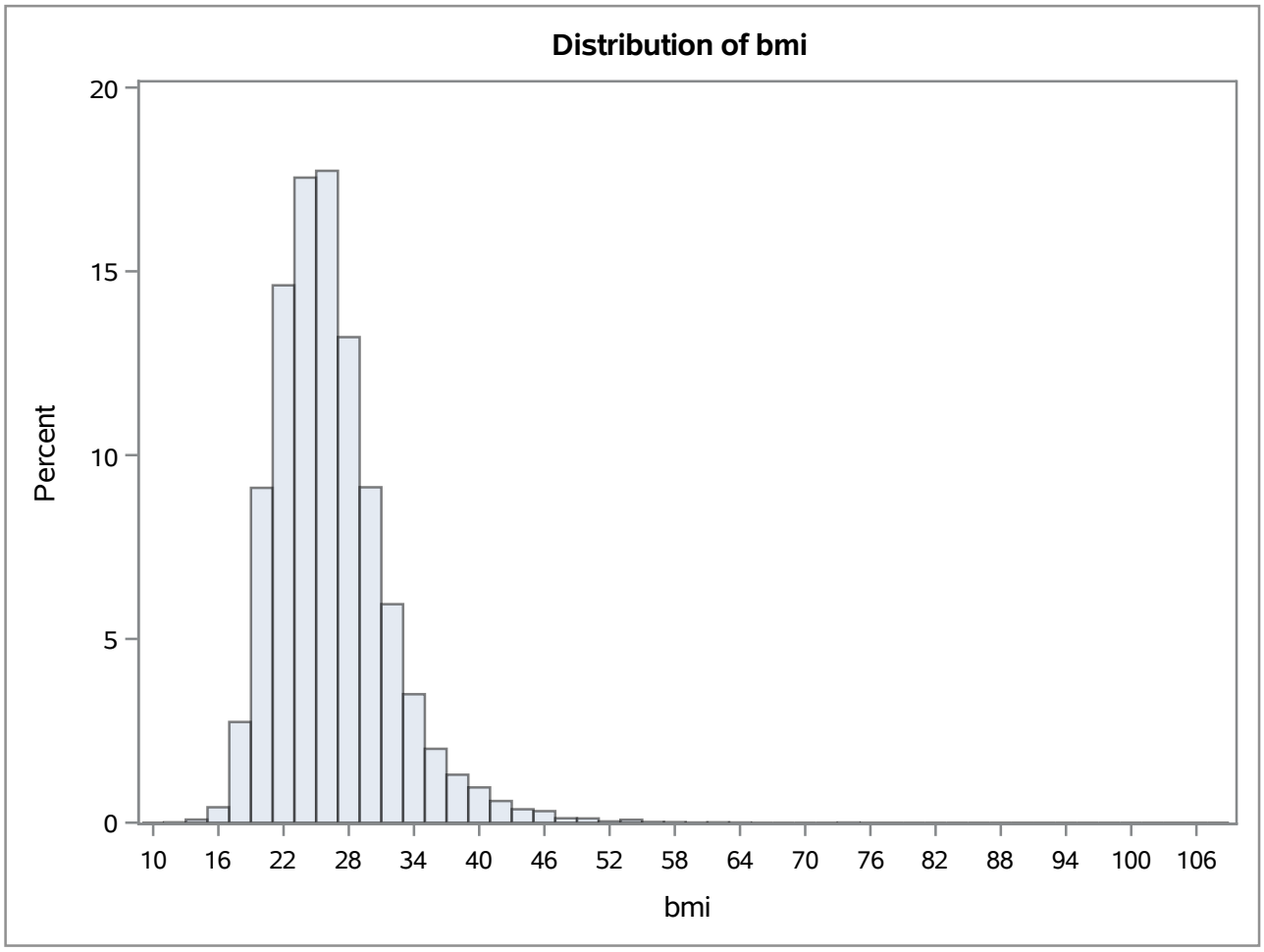
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	712.9717	**Pr > |t|**	<.0001
Sign	**M**	10000	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.0001E8	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	73.0907
99%	43.5553
95%	35.9509
90%	32.8802
75% Q3	28.8862
50% Median	25.6035
25% Q1	22.7077
10%	20.5957
5%	19.4835
1%	17.7123
0% Min	12.4005

The UNIVARIATE Procedure
Variable: bmi

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
12.4005	15887	60.2217	443
12.7495	3905	61.5733	17968
13.3872	19127	62.6420	10060
13.5565	8698	64.1892	11392
13.8104	2493	73.0907	2284

The UNIVARIATE Procedure



The UNIVARIATE Procedure
Variable: age

Moments			
N	20000	**Sum Weights**	20000
Mean	45.06825	**Sum Observations**	901365
Std Deviation	17.1926895	**Variance**	295.588571
Skewness	0.45170032	**Kurtosis**	-0.6548259
Uncorrected SS	46534419	**Corrected SS**	5911475.84
Coeff Variation	38.1481186	**Std Error Mean**	0.12157067

Basic Statistical Measures			
Location		**Variability**	
Mean	45.06825	**Std Deviation**	17.19269
Median	43.00000	**Variance**	295.58857
Mode	40.00000	**Range**	81.00000
		Interquartile Range	26.00000

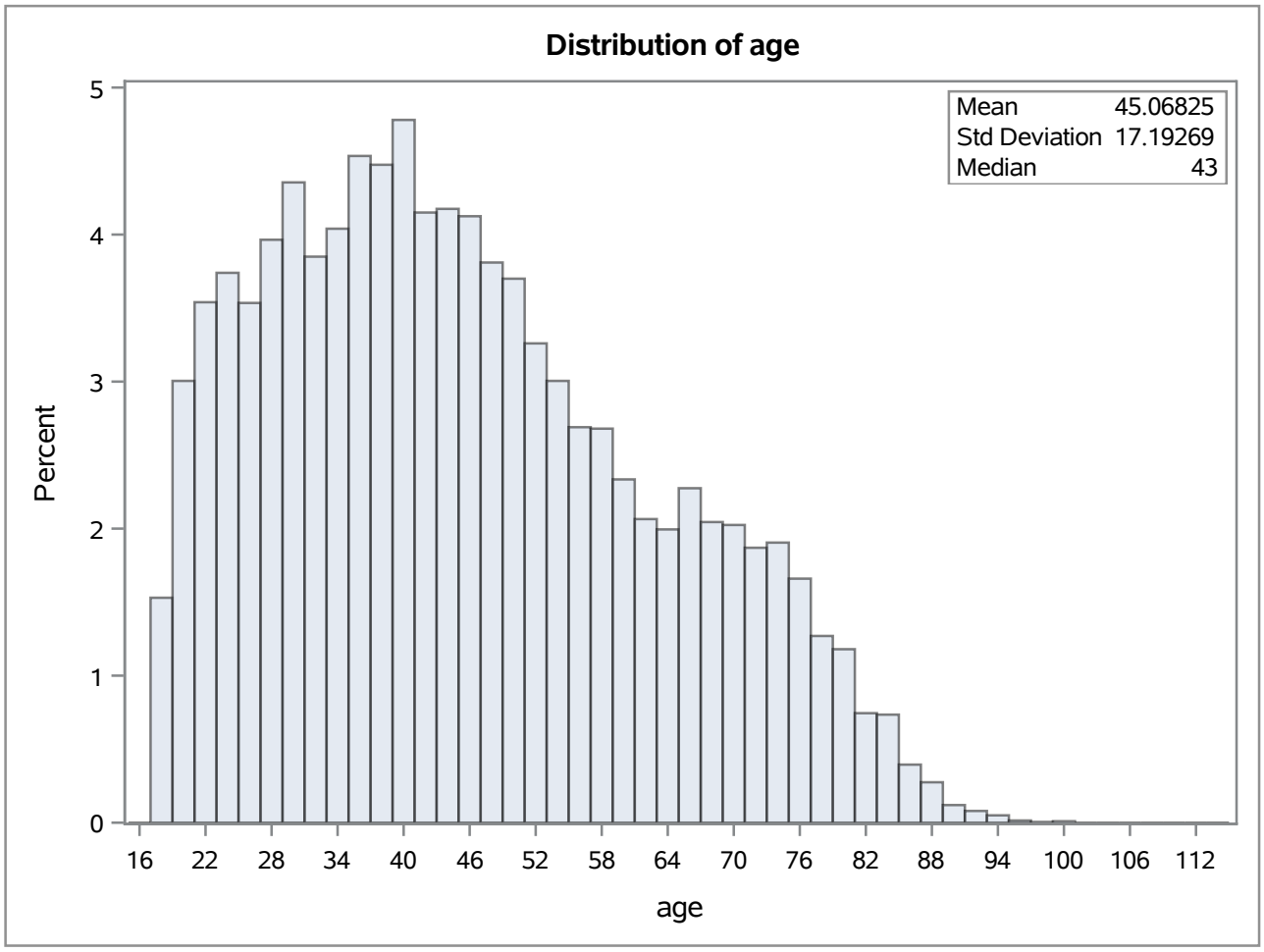
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	370.7165	**Pr > |t|**	<.0001
Sign	**M**	10000	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.0001E8	**Pr >= |S|**	<.0001

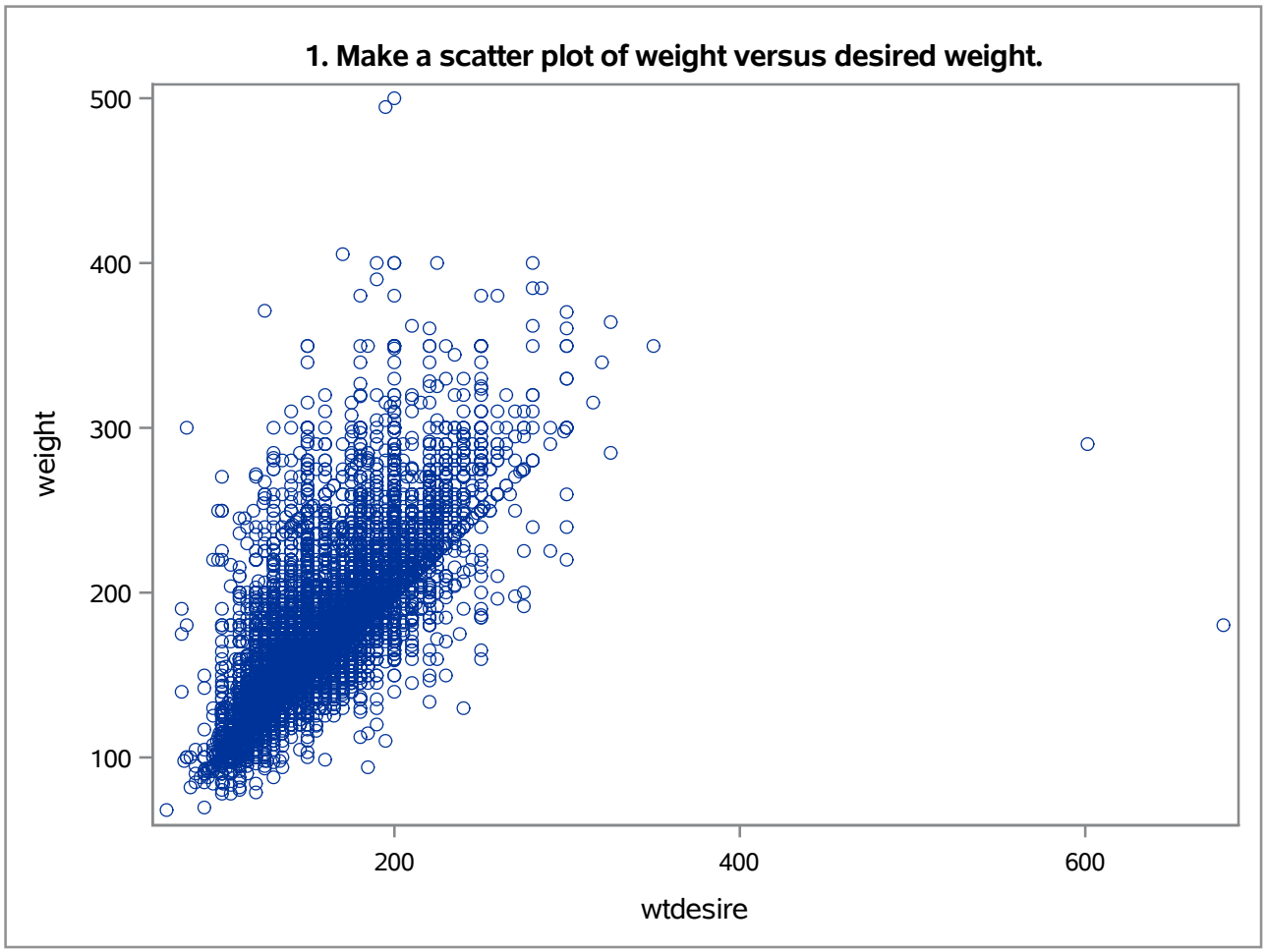
Quantiles (Definition 5)	
Level	**Quantile**
100% Max	99
99%	84
95%	76
90%	71
75% Q3	57
50% Median	43
25% Q1	31
10%	24
5%	21
1%	18
0% Min	18

The UNIVARIATE Procedure
Variable: age

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
18	19950	95	8360
18	19941	96	8855
18	19932	97	5333
18	19869	99	479
18	19827	99	3428

The UNIVARIATE Procedure




1. Make a scatter plot of weight versus desired weight.
500
400
300
200
100
weight
200
400
600
wtdesire

2. Create this new variable by subtracting the two columns in the new data set andassigning them to a new object named wdiff.

Description
The two variables increase the same way.

2. Create this new variable by subtracting the two columns in the new data set andassigning them to a new object named wdiff.

Obs	genhlth	exerany	hlthplan	smoke100	height	weight	wtdesire	age	gender	wdiff
1	good	0	1	1	64	125	115	33	f	-10
2	good	1	1	1	60	105	105	49	f	0
3	good	1	1	0	66	132	124	42	f	-8
4	very good	0	1	0	61	150	130	55	f	-20
5	very good	1	1	0	64	114	114	55	f	0
6	good	0	1	1	65	150	130	27	f	-20
7	very good	1	1	0	67	125	120	33	f	-5
8	very good	1	1	0	69	200	150	48	f	-50
9	good	1	1	1	65	160	140	54	f	-20
10	very good	0	0	1	64	105	120	27	f	15

4. Describe the distribution of wdiff in terms of its center, shape, and spread, including any plots you use. What does this tell us about how people feel about their current weight?

Type	wdiff_equals_0	wdiff_positive_negative
Both wtdesire and weight are numerical. Thus, the difference of the two variables is also numerical.	If wdiff is 0, then the person has the weight equal to the desired.	If wdiff is positive then the weight is smaller than the desired.Also if wdiff is negative then the weight is larger than the desired.

4. Describe the distribution of wdiff in terms of its center, shape, and spread, including any plots you use. What does this tell us about how people feel about their current weight?

The UNIVARIATE Procedure
Variable: wdiff

Moments			
N	20000	**Sum Weights**	20000
Mean	-14.5891	**Sum Observations**	-291782
Std Deviation	24.0458556	**Variance**	578.203171
Skewness	-1.4477403	**Kurtosis**	21.6223956
Uncorrected SS	15820322	**Corrected SS**	11563485.2
Coeff Variation	-164.82069	**Std Error Mean**	0.17002988

Basic Statistical Measures			
Location		**Variability**	
Mean	-14.5891	**Std Deviation**	24.04586
Median	-10.0000	**Variance**	578.20317
Mode	0.0000	**Range**	800.00000
		Interquartile Range	21.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	-85.8032	**Pr > |t|**	<.0001
Sign	**M**	-5572	**Pr >= |M|**	<.0001
Signed Rank	**S**	-4.297E7	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	500.0
99%	25.0
95%	10.0
90%	0.0
75% Q3	0.0
50% Median	-10.0
25% Q1	-21.0
10%	-40.0
5%	-59.5
1%	-100.0
0% Min	-300.0

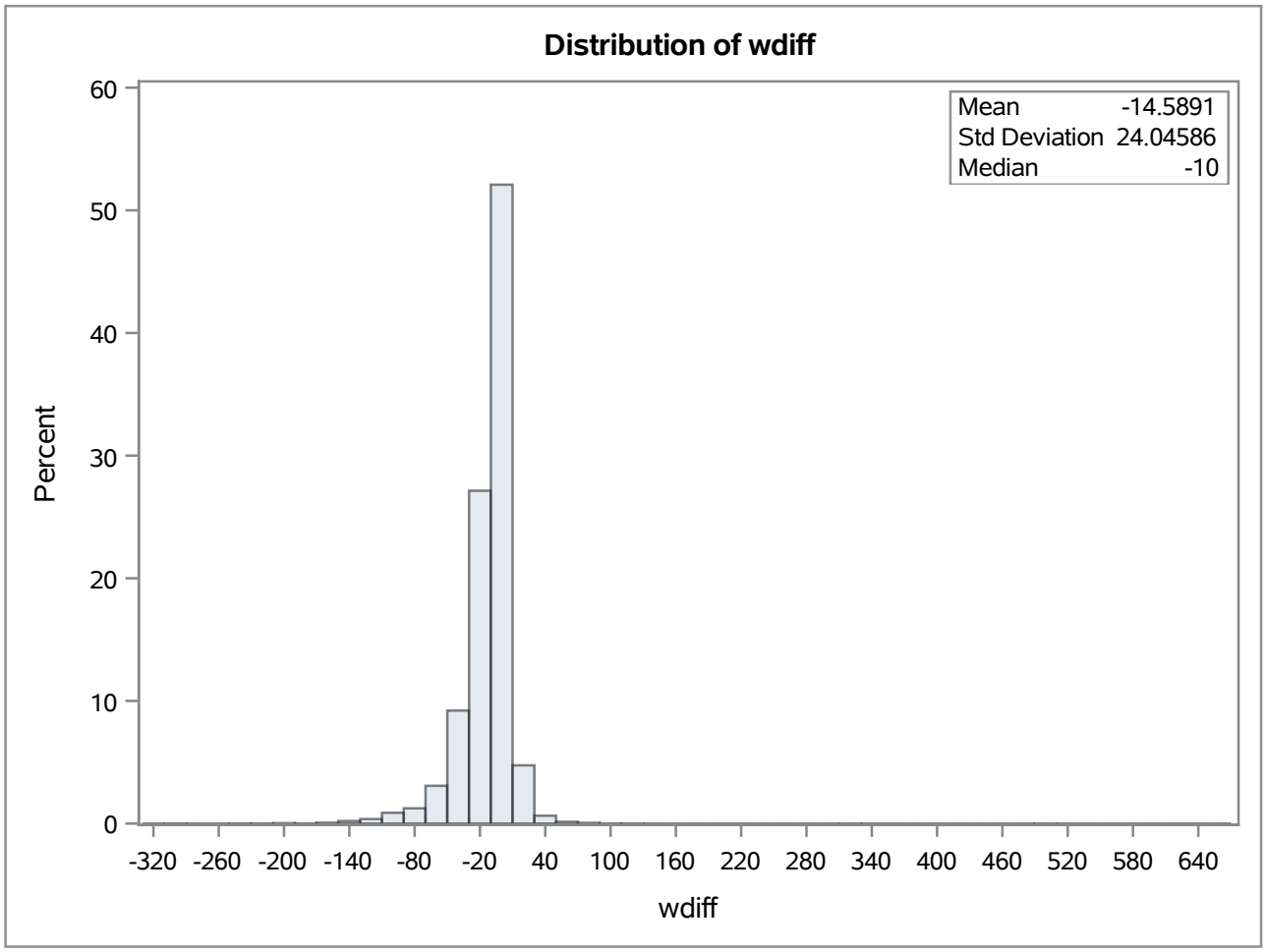
4. Describe the distribution of wdiff in terms of its center, shape, and spread, including any plots you use. What does this tell us about how people feel about their current weight?

The UNIVARIATE Procedure
Variable: wdiff

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-300	11392	90	14949
-300	2284	91	15887
-246	2354	110	13353
-235	17968	311	15296
-220	4628	500	18534

4. Describe the distribution of wdiff in terms of its center, shape, and spread, including any plots you use. What does this tell us about how people feel about their current weight?

The UNIVARIATE Procedure



5. Using numerical summaries and a side-by-side box plot, determine whether men tend to view theirweight differently than women.

Answer
The distribution of wdiff tells us that the people have weight larger than the desired because the wdiff is negative.

5. Using numerical summaries and a side-by-side box plot, determine whether men tend to view theirweight differently than women.

The UNIVARIATE Procedure
Variable: wdiff

gender=f

Moments			
N	10431	**Sum Weights**	10431
Mean	-18.151184	**Sum Observations**	-189335
Std Deviation	23.9971327	**Variance**	575.862376
Skewness	-2.2611475	**Kurtosis**	9.22617222
Uncorrected SS	9442899	**Corrected SS**	6006244.58
Coeff Variation	-132.20698	**Std Error Mean**	0.23496132

Basic Statistical Measures			
Location		**Variability**	
Mean	-18.1512	**Std Deviation**	23.99713
Median	-10.0000	**Variance**	575.86238
Mode	0.0000	**Range**	383.00000
		Interquartile Range	27.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	-77.2518	**Pr > |t|**	<.0001
Sign	**M**	-3533.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	-1.486E7	**Pr >= |S|**	<.0001

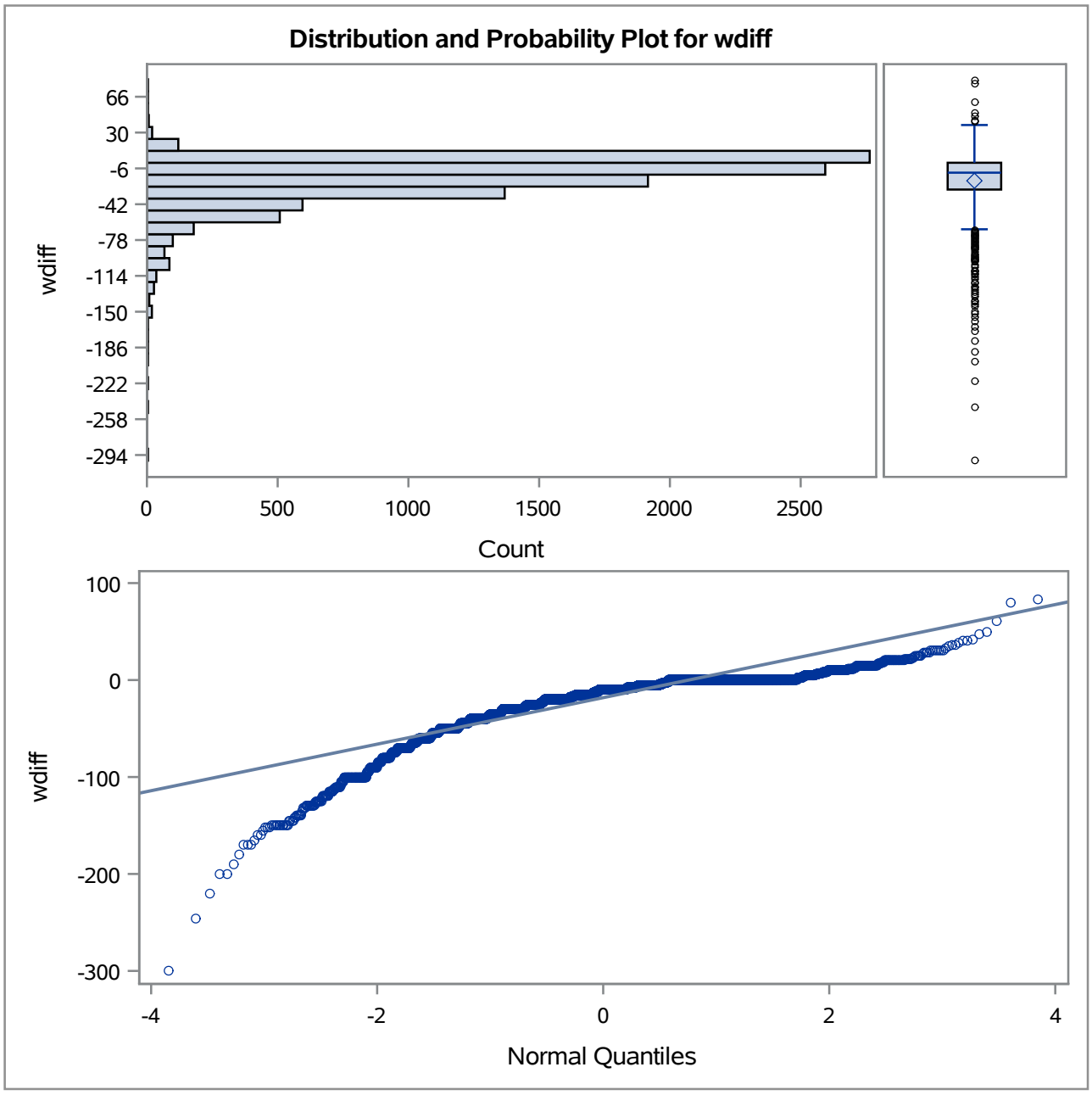
Quantiles (Definition 5)	
Level	**Quantile**
100% Max	83
99%	15
95%	0
90%	0
75% Q3	0
50% Median	-10
25% Q1	-27
10%	-48
5%	-63
1%	-108
0% Min	-300

5. Using numerical summaries and a side-by-side box plot, determine whether men tend to view theirweight differently than women.

The UNIVARIATE Procedure
Variable: wdiff

gender=f

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-300	2284	47	2294
-246	2354	50	2431
-220	4628	61	2076
-200	10060	80	5762
-200	784	83	7240



5. Using numerical summaries and a side-by-side box plot, determine whether men tend to view theirweight differently than women.

The UNIVARIATE Procedure
Variable: wdiff

gender=m

Moments			
N	9569	**Sum Weights**	9569
Mean	-10.706134	**Sum Observations**	-102447
Std Deviation	23.4926242	**Variance**	551.903392
Skewness	-0.5973152	**Kurtosis**	37.4730657
Uncorrected SS	6377423	**Corrected SS**	5280611.65
Coeff Variation	-219.43143	**Std Error Mean**	0.24015866

Basic Statistical Measures			
Location		**Variability**	
Mean	-10.7061	**Std Deviation**	23.49262
Median	-5.0000	**Variance**	551.90339
Mode	0.0000	**Range**	800.00000
		Interquartile Range	20.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	-44.5794	**Pr > |t|**	<.0001
Sign	**M**	-2038.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	-7097976	**Pr >= |S|**	<.0001

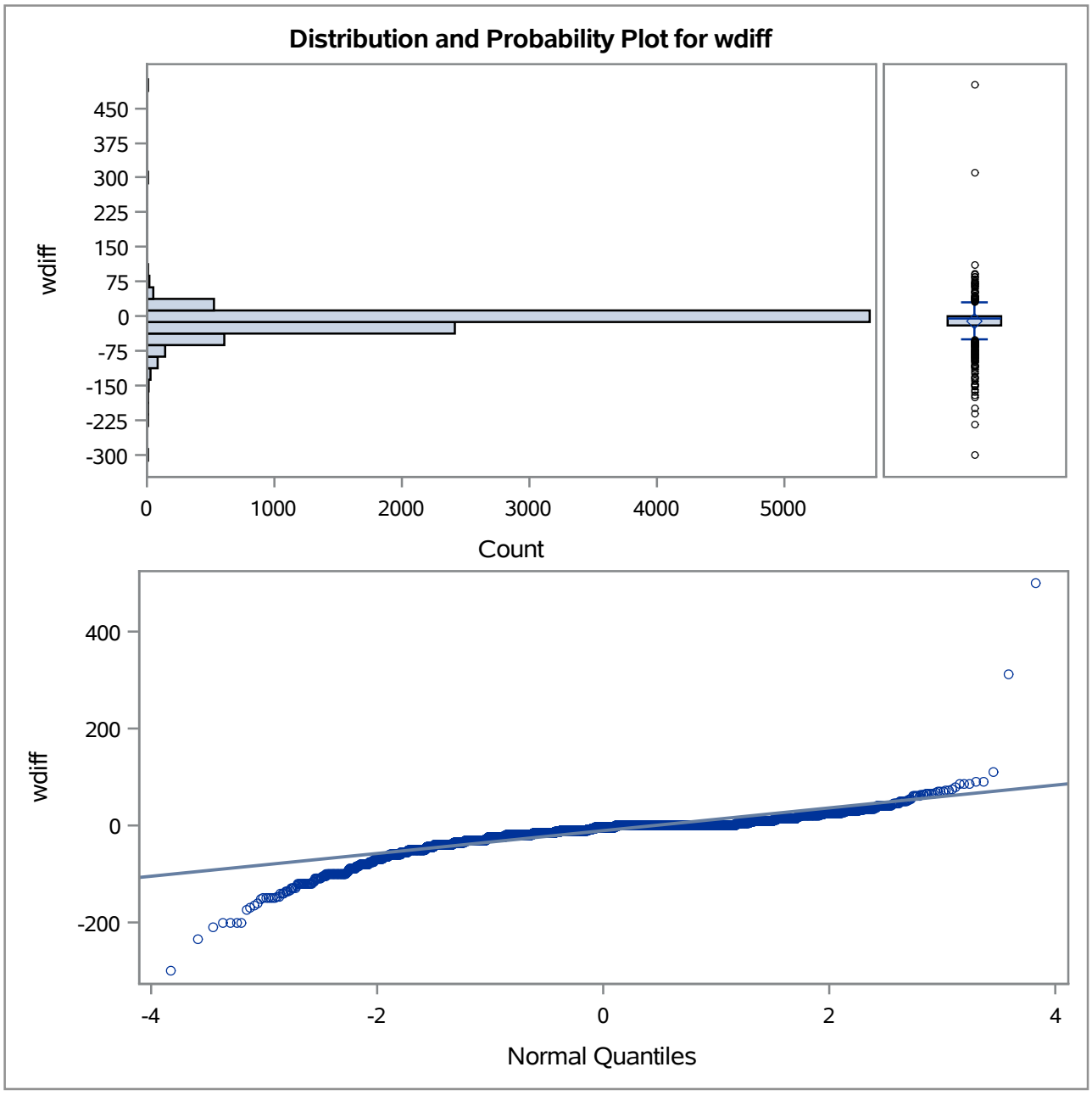
Quantiles (Definition 5)	
Level	**Quantile**
100% Max	500
99%	35
95%	15
90%	6
75% Q3	0
50% Median	-5
25% Q1	-20
10%	-35
5%	-50
1%	-100
0% Min	-300

5. Using numerical summaries and a side-by-side box plot, determine whether men tend to view theirweight differently than women.

The UNIVARIATE Procedure
Variable: wdiff

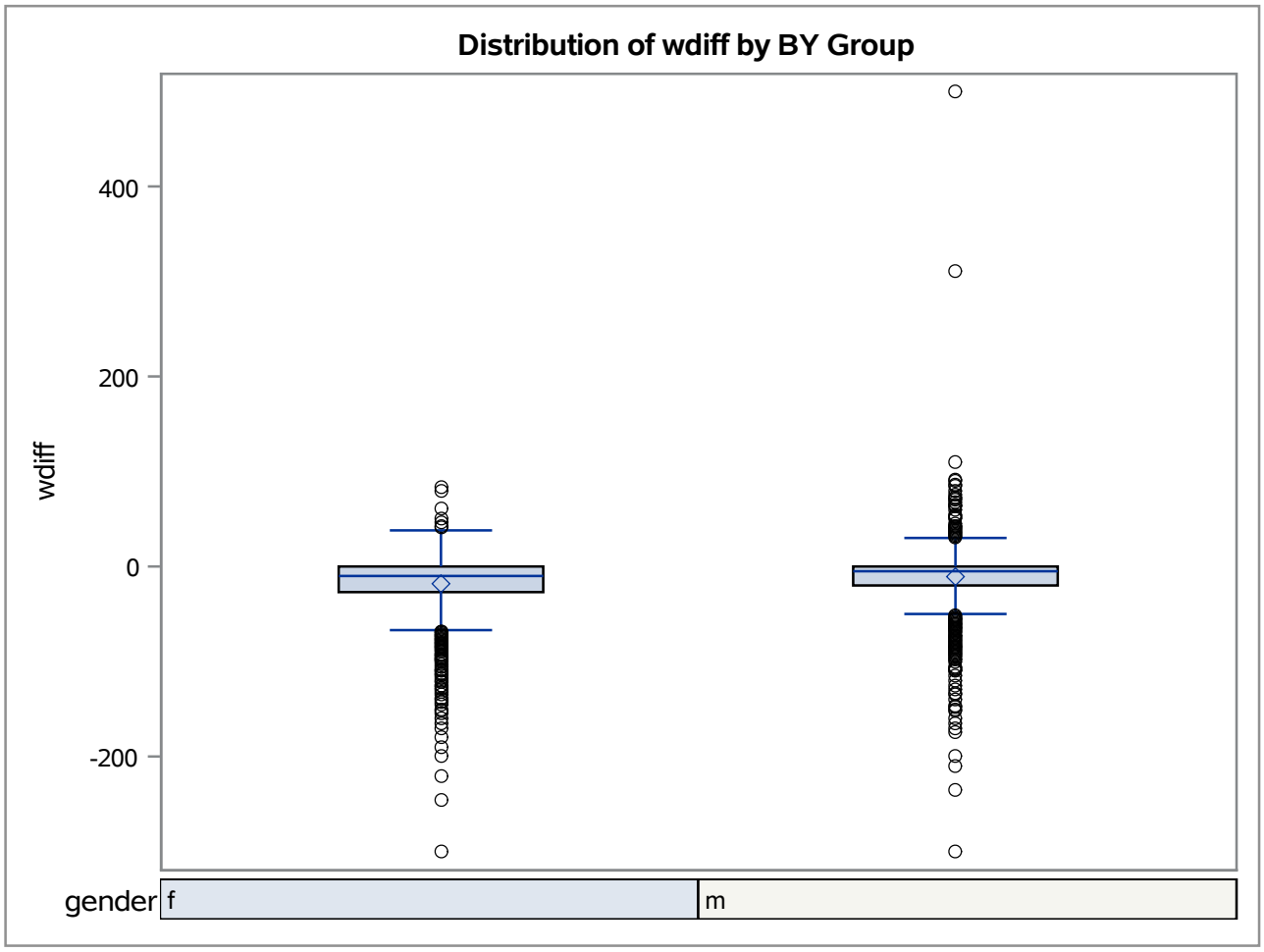
gender=m

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-300	11392	90	14949
-235	17968	91	15887
-210	11867	110	13353
-200	18555	311	15296
-200	17510	500	18534



5. Using numerical summaries and a side-by-side box plot, determine whether men tend to view theirweight differently than women.

The UNIVARIATE Procedure



6. Now it's time to get creative. Find the mean and standard deviation of weight and determine whatproportion of the weights is within one standard deviation of the mean.

The UNIVARIATE Procedure
Variable: weight

Moments			
N	20000	**Sum Weights**	20000
Mean	169.68295	**Sum Observations**	3393659
Std Deviation	40.08097	**Variance**	1606.48415
Skewness	0.95572799	**Kurtosis**	1.99615504
Uncorrected SS	607974147	**Corrected SS**	32128076.6
Coeff Variation	23.6210945	**Std Error Mean**	0.28341526

Basic Statistical Measures			
Location		**Variability**	
Mean	169.6830	**Std Deviation**	40.08097
Median	165.0000	**Variance**	1606
Mode	160.0000	**Range**	432.00000
		Interquartile Range	50.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	598.7079	**Pr > |t|**	<.0001
Sign	**M**	10000	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.0001E8	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	500
99%	290
95%	240
90%	220
75% Q3	190
50% Median	165
25% Q1	140
10%	124
5%	115
1%	100
0% Min	68

6. Now it's time to get creative. Find the mean and standard deviation of weight and determine whatproportion of the weights is within one standard deviation of the mean.

The UNIVARIATE Procedure
Variable: weight

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
68	9739	400	11725
70	8592	400	11867
78	19127	405	17968
78	5860	495	2284
79	3905	500	11392