Solutions for Advanced Statistics Assignment

1.

	Mean	Median	Variance	Standard Deviation	Degrees of freedom
M	176.75	180	64.92	8.06	3
F	162.25	162.5	74.25	8.62	3
Population	169.5	169.5	119.714	10.941	7

2.

	Response		
Treatment	no	yes	Tot
A	85	37	122
B	32	32	64
Tot	117	69	186

There were 122 patients in group A, 37 of them responded to treatment. In total, 186 patients were treated, and 69 of them responded.
Now we compute the appropriate percentages:

	Response	
Treatment	no	yes
A	85/122=69.7%	37/122=30.3%
B	32/64=50.0%	32/64=50.0%
Tot	117/186=62.9%	69/186=37.1%

The best treatment seems to be treatment B.
Since 50/30.3=1.65
So response of treatment B is 65 % higher than treatment A.
Among responders, the percent of those who got treatment B was: 32/69=46,4%.
This percentage is obtained by considering the distribution of Treatment conditional on Response equal to Yes.

3.
(54*25)+(58*62)=4946
4946/(25+62)=56.85 is the overall mean

4.

Mean	64.6625
Standard Error	2.8987643611
Mode	
Median	63.15
First Quartile	60.45
Third Quartile	66.4
Variance	67.2226785714
Standard Deviation	8.1989437473
Kurtosis	2.1759895056
Skewness	1.2464164793
Range	27
Minimum	54.5
Maximum	81.5
Sum	517.3
Count	8

mean=64.6625
median=63.15

There is not much difference between mean and median. And thus, the data is not highly skewed.

5.

a.

- 1 out of 4 was younger than 27 years: this is the definition of first quartile, ¼=25% of observed values was lower than Q1=27
- 1 out of 4 was older than 59 years: similarly, this is the definition of third quartile, ¾=75% of observations was lower than Q3, and the other 25% was larger than Q3=59
- 2 out of 4 could have three possibilities:
  - "between 0 years and the median 41 years", but also
  - "between Q1 and Q3" and also
  - "between the median 41 and the maximum age"
- Half of them was more than 41 years old: this comes from the definition of the Median.

b.

First we notice that the mean is 42 and it is very close to the median,
in fact their distance (equal to 1) is small (1/12) compared to the standard deviation. Thus the observed distribution is rather symmetric.
But the Normal is not the only kind of symmetric distribution; we can go further by looking at the quartiles.
In a Normal curve the first and third quartile should be at a specific distance from the mean, given by 0.67 times the standard deviation, thus 0.67·12=8. Thus if the observed distribution was approximately Normal the first and third quartiles were expected to be (42+8=50 and 42-8=34).
Our observed quartiles are instead 27 and 59, rather distant from the ones of a Normal distribution with the same mean and standard deviation. Thus, our distribution is NOT Normal-shaped; it was symmetric but not bell-shaped; it could have been a distribution with very high tails and few observations in the middle, possibly a distribution with two modes..

c.

Neither the mean nor the median are good indexes to describe the distribution.

6.

no. of bulbs =1000-(1000*(2.5/100))=975

$$\frac{(1000-975)*\sqrt{n}}{125}=\frac{25*\sqrt{n}}{125}$$

$\frac{\sqrt{n}}{5}\leq 1.29$ (Refer to figure below for 90% confidence 0.9015 in z-table)

n=41.6025

The sample size is 41.

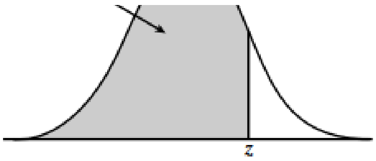

Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981

7.
Let p and P denote the population and sample proportions of shoppers not making a purchase.

H_0: p=P
H_1: p>P, since p=0.7 and P=0.6

One tailed test is to be used.
Let LOS be 5%. Therefore, z_α=1.645.

$$z=\frac{p-P}{\sqrt{\frac{PQ}{n}}}=\frac{0.7-0.6}{\sqrt{\frac{0.6\,x\,0.4}{50}}}=1.443$$

$|z|< z_{\alpha}$.
Thus the null hypothesis is accepted and the results are consistent with the claims of the salesman.

8.
a.)
μ+2σ=72+16=88
μ-2σ=72-16=56

(56,88)

b.)

$\mu+3\sigma=72+24=96$

$\mu-3\sigma=72-24=48$

(48,96)

c.)

50 % of observations are covered under Q3-Q1

$\mu+0.67*\sigma=66.64$

$\mu-0.67*\sigma=77.36$

9.

Here , credit history and age, credit history and gender, gender and age,job age are strongly correlated with each other as correlation value is above 0.75 (irrespective of sign). Covariance just tells us the sign of relationship which is of not much use in this question as correlation values also give the direction.