

Homework 3, due in Lecture Tuesday, Sept 27th ---- Solutions

(assignment for these solutions can be found on the last page)

1. Separately for males and females, assess the normality of the distribution of birth weights, using normal scores plots, histograms, and side-by-side boxplots. Describe all important features of each sex's weight distribution, and explain how you can see the features in all three displays. You probably will want to "unstack" the data into separate columns in order to get the normal scores. How do the distributions of weight for males compare to those for females?

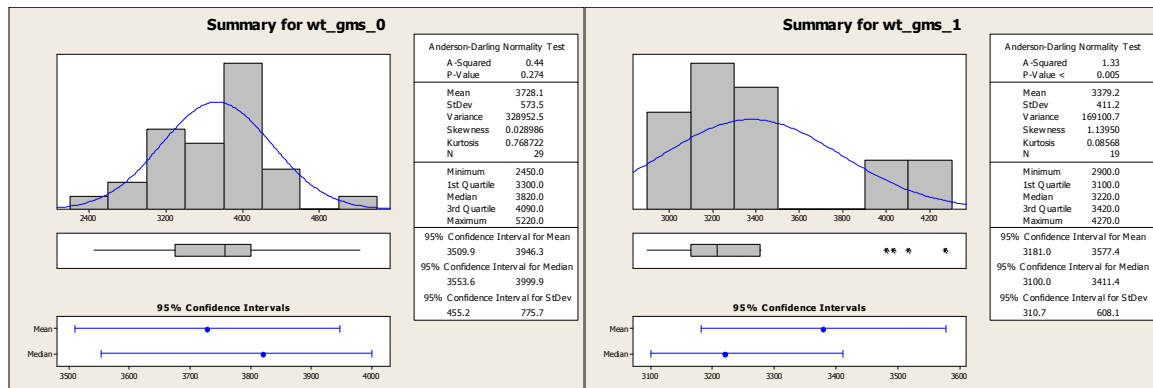
Using the normal scores plots below (second set of plots) we notice the boy's weights are enough in a line to consider normal, while the girl's weights are not. This is supported by the p-value associated with a test for normality, also reported in the plot.

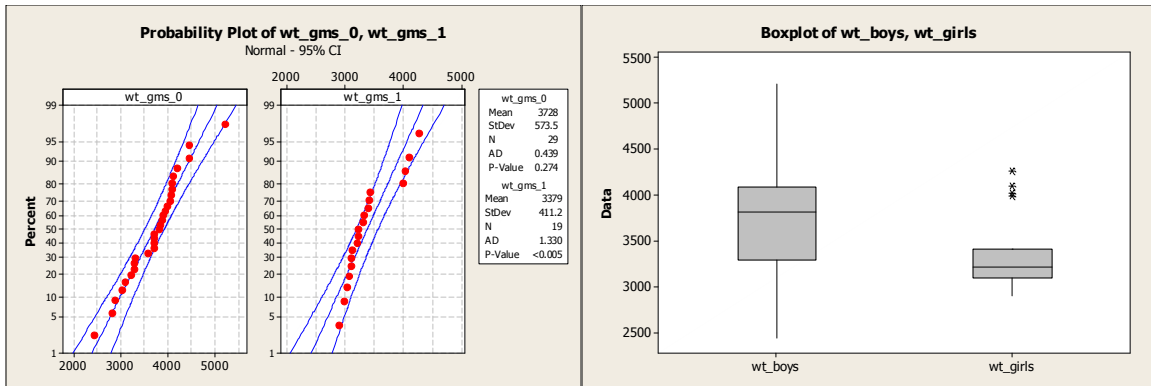
The histograms below show the boy's weights roughly normal, that is, symmetric, unimodal, no outliers. The box plot is roughly symmetric, no outliers. The girl's weights appear in two groups, most of them in a low group, and some in a high group. The box plot indicates this upper group as outliers.

The descriptive statistics below indicate that boys are heavier than girls (mean and median), but are also more variable (stdev is bigger, and the min is smaller and the max larger for boys than girls).

Descriptive Statistics: wt_gms_0, wt_gms_1

| Variable | N | Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
|----------|----|--------|-------|---------|--------|--------|--------|---------|
| wt_gms_0 | 29 | 3728 | 574 | 2450 | 3300 | 3820 | 4090 | 5220 |
| wt_gms_1 | 19 | 3379.2 | 411.2 | 2900.0 | 3100.0 | 3220.0 | 3420.0 | 4270.0 |



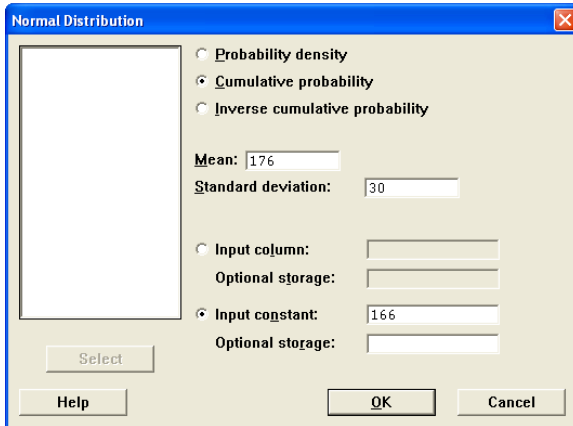


2. Do problem 5.15 in SW, p. 164.

Serum cholesterol levels are Normal(mean=176, std=30).

a. $\Pr(166 < Y < 186) = 0.630559 - 0.369441 = \mathbf{0.261118}$.

This result can be seen from this screenshot and two cumulative distribution calculations:



Cumulative Distribution Function

Normal with mean = 176 and standard deviation = 30

| | |
|-----|-------------|
| x | P(X <= x) |
| 166 | 0.369441 |

Cumulative Distribution Function

Normal with mean = 176 and standard deviation = 30

| | |
|-----|-------------|
| x | P(X <= x) |
| 186 | 0.630559 |

b. The sampling distribution for samples of size 9 from this normal population has the same mean but divides the standard deviation by the square root of the sample size. Thus, it is
Normal(mean=176, std=30/sqrt(9)=10).

Performing the same calculation with this mean and standard deviation we have $\Pr(166 < \bar{Y} < 186) = 0.841345 - 0.158655 = \mathbf{0.68269}$.

Note that this is 1 standard deviation out from the mean in either direction, so roughly 68% of the area is there.

c. **Same as (b).**

3. Do problem 5.16 in SW, p. 164

Original distribution is Normal(mean=3000, stdev=400). We want $\Pr(2900 < \bar{Y} < 3100)$

a. $n=15$ gives Normal(3000, $400/\sqrt{15}=103.28$)

$\Pr(2900 < \bar{Y} < 3100) = 0.833538 - 0.166462 = \mathbf{0.667076}$

b. $n=60$ gives Normal(3000, $400/\sqrt{60}=51.64$)

$\Pr(2900 < \bar{Y} < 3100) = 0.973596 - 0.026404 = \mathbf{0.947192}$

c. **As n increases, $\Pr(\bar{E})$ increases**, which is why we always want to use larger sample sizes when possible.

4. Do problem 5.34 in SW, p. 174

We want to calculate the area with .03 of .12, so from .09 to .15, that is $\Pr(.09 < p < .15)$ using the normal approximation.

a. Using Theorem 5.2a, the normal approximation for the binomial parameter p for a sample size $n=100$ is Normal(.12, $\sqrt{.12(1-.12)/100} = 0.0325$)

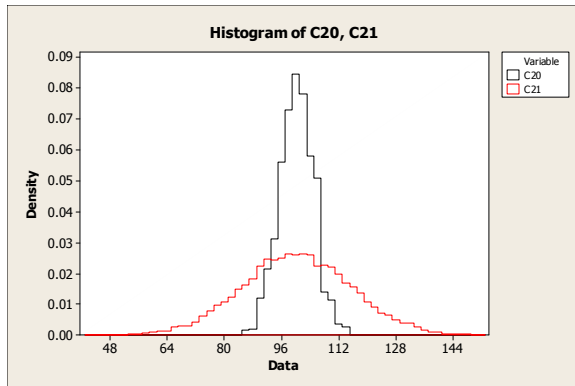
$\Pr(.09 < p < .15) = 0.822016 - 0.177984 = \mathbf{0.644032}$

b. With $n=200$ is Normal(.12, $\sqrt{.12(1-.12)/200} = 0.0230$)

$\Pr(.09 < p < .15) = 0.903942 - 0.096058 = \mathbf{0.807884}$

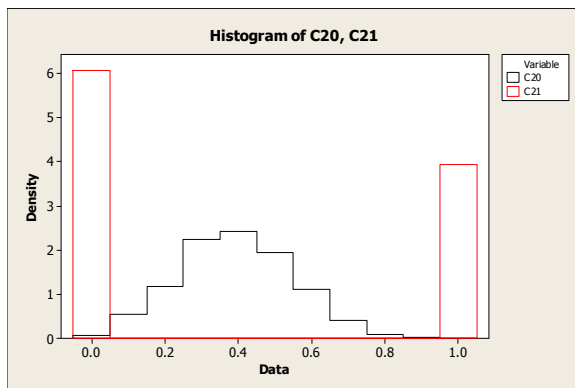
5. Display the results of a simulation of the sampling distribution of the sample mean when $n=10$ sampling from a normal distribution with population mean 100 and population standard deviation 15. Show histograms of both the original population and the sampling distribution of the sample mean overlaid on the same graph.

Lab 4 describes the process for doing this. I create 1000 rows and 10 columns from a Normal(100,15) distribution. Then I perform row means across the 10 columns to produce a column with means. Then I stack the 10 columns into one. At last, I create a histogram with both the mean column and stacked column. Below is the result. C20 (the tall histogram) is the means, and C21 (wider) is the original distribution. Both of these distributions are normal, but the distribution of means has a small standard deviation.



6. Repeat problem 5 for the sampling distribution of the sample proportion when sampling from a binary population with 40% S's and 60% F's.

Also described in Lab 4. I create 1000 rows and 10 columns from a Bernoulli(0.4) distribution. Then I perform row means across the 10 columns to produce a column with means. Then I stack the 10 columns into one. At last, I create a histogram with both the mean column and stacked column. Below is the result. C20 (the tall histogram) is the means, and C21 (wider) is the original distribution. The original distribution is clearly not normal (two towers at 0 and 1), but the distribution of means looks close to normal.



Stat 538 - Biostatistics I

Homework 3, due in Lecture Tuesday, Sept 27.

1. The following data were collected from 48 women who were at least 40 years old when they gave birth to their first child. The data concern the gestation period of that pregnancy, and related variables on the child and mother.

The columns are, from left to right:

- 1) ID
- 2) The child's gestation period, in weeks
- 3) Sex of the child (0=Male, 1=Female)
- 4) Birth Weight of child, in grams
- 5) Number of cigarettes smoked per day (on average) by the mother
- 6) Height of mother in cm
- 7) Weight of mother in kilograms at first prenatal visit
- 8) Weight of mother in kilograms at final prenatal visit

Separately for males and females, assess the normality of the distribution of birth weights, using normal scores plots, histograms, and side-by side boxplots. Describe all important features of each sex's weight distribution, and explain how you can see the features in all three displays. You probably will want to "unstack" the data into separate columns in order to get the normal scores. How do the distributions of weight for males compare to those for females?

Save this worksheet - we will use the data later.

DATA

```
----  
 1 36 0 3300 0 160.0 67.3 82.7  
 2 38 0 3300 60 167.6 52.7 76.0  
 3 38 0 4100 20 167.6 64.2 79.6  
 4 38 1 2900 10 163.9 72.7 95.8  
 5 39 0 2820 0 161.3 50.0 63.3  
 6 39 0 3040 0 158.8 49.1 61.5  
 7 39 0 4120 0 160.0 57.7 73.5  
 8 39 0 4200 0 174.0 68.0 86.8  
 9 39 1 3100 0 171.5 67.3 85.6  
10 39 1 3330 0 160.0 74.0 90.5  
11 39 1 3410 0 165.1 55.9 70.7  
12 39 1 3420 0 162.6 52.3 66.0  
13 40 0 2450 20 167.6 61.4 72.5  
14 40 0 2885 0 167.7 60.0 78.6  
15 40 0 3235 0 170.2 50.0 65.5  
16 40 0 3320 0 165.1 63.6 80.2  
17 40 0 3600 0 165.1 53.2 68.7  
18 40 0 3720 0 165.0 57.7 74.4  
19 40 0 3720 0 172.7 61.4 80.0  
20 40 0 3820 0 175.3 60.8 78.1  
21 40 0 3840 0 167.0 60.5 83.9  
22 40 0 3880 0 156.2 57.3 73.7  
23 40 0 3960 0 157.5 52.7 68.2  
24 40 0 4465 0 157.5 51.4 66.4  
25 40 1 2980 0 160.0 47.7 55.2  
26 40 1 3040 0 162.0 49.0 60.3  
27 40 1 3060 20 157.5 61.0 75.0  
28 40 1 3100 0 170.2 55.5 64.6  
29 40 1 3120 0 160.3 56.8 75.4  
30 40 1 3205 0 172.7 58.2 75.5  
31 40 1 3220 0 170.0 64.6 86.0  
32 40 1 4100 40 167.0 67.0 85.0  
33 41 0 3100 0 168.9 61.4 69.2  
34 41 0 3720 0 170.2 57.7 67.7  
35 41 0 3720 20 170.2 57.7 80.5
```

Stat 538 – Biostatistics I
Homework 3 solutions

| | | | | | | | |
|----|----|---|------|----|-------|------|-------|
| 36 | 41 | 0 | 3900 | 0 | 167.0 | 68.0 | 85.4 |
| 37 | 41 | 0 | 3990 | 0 | 165.1 | 52.3 | 71.2 |
| 38 | 41 | 0 | 4050 | 0 | 167.6 | 61.0 | 78.5 |
| 39 | 41 | 0 | 4080 | 0 | 162.6 | 59.1 | 83.1 |
| 40 | 41 | 0 | 4100 | 0 | 165.1 | 60.5 | 86.5 |
| 41 | 41 | 0 | 4460 | 20 | 165.1 | 56.8 | 88.0 |
| 42 | 41 | 0 | 5220 | 0 | 157.5 | 56.8 | 68.2 |
| 43 | 41 | 1 | 3300 | 40 | 162.6 | 74.1 | 89.7 |
| 44 | 41 | 1 | 3400 | 0 | 172.7 | 71.4 | 87.8 |
| 45 | 41 | 1 | 4000 | 0 | 165.1 | 90.0 | 100.8 |
| 46 | 41 | 1 | 4030 | 0 | 166.0 | 63.0 | 95.3 |
| 47 | 43 | 1 | 3220 | 0 | 166.4 | 60.9 | 72.0 |
| 48 | 43 | 1 | 4270 | 0 | 162.6 | 54.5 | 70.3 |

You may use either the normal tables or Minitab for calculating probabilities with the following three problems from SW.

2. Do problem 5.15 in SW, p. 164.
3. Do problem 5.16 in SW, p. 164
4. Do problem 5.34 in SW, p. 174
5. Display the results of a simulation of the sampling distribution of the sample mean when $n=10$ sampling from a normal distribution with population mean 100 and population standard deviation 15. Show histograms of both the original population and the sampling distribution of the sample mean overlaid on the same graph.
6. Repeat problem 5 for the sampling distribution of the sample proportion when sampling from a binary population with 40% S's and 60% F's.