Statistics 3540
Assignment #1: Solutions – 30 POINTS

1. Problem on the sampling distribution of $\bar{x}$.

   NOTE: No computer output is required in the solutions. Also, not everyone will have the same results, since the values are randomly generated.

   (a) From the output, we see that $\mu = 16$ and $\sigma = 5$.
   (b) No answer needed.
   (c) The histogram that I was shown did not have much of a shape at all (results will vary).
   (d) From my output, the mean of my $\bar{x}$ values was 17.48, and the standard deviation was 1.08 (again, everyone will have different answers).
   The Central Limit Theorem states that we'd expect the mean to be $\mu = 16$ and the standard deviation to be $\sigma/\sqrt{n} = 5/\sqrt{20} = 1.12$. For my computer results, the mean isn't that close to 16, although the standard deviation isn't that far from what the theorem states.
   (e) No response needed.
   (f) My histogram, based on 1000 samples, appeared symmetric (again, results will vary).
   (g) From the output, the mean is 16.06 and the standard deviation is 1.16. Although the mean is much closer to what the Central Limit Theorem predicts, the standard deviation hasn't improved that much. I guess that goes to show the theory tells us what to expect will happen, and not what is guaranteed to happen.

---

2. Interpreting confidence intervals.

   (a) Some of my 60 confidence intervals (which will be different than yours) are below (you will not lose points if you didn't submit this portion of output):

```
One-Sample Z: C1, C2, C3, C4, C5, C6, C7, C8, ...

The assumed standard deviation = 3


Variable   N     Mean    StDev  SE Mean          95% CI
C1        50  9.59721  2.94393  0.42426  (8.76567, 10.42875)
C2        50  10.0109   3.0413   0.4243  ( 9.1793,  10.8424)
C3        50  10.2672   2.8788   0.4243  ( 9.4357,  11.0988)
 .
 .
 .
C58       50  9.82494  2.80372  0.42426  (8.99340, 10.65648)
C59       50  9.51952  2.94080  0.42426  (8.68798, 10.35107)
C60       50  10.1108   2.7992   0.4243  ( 9.2793,  10.9424)
```

(b) 58 of my confidence intervals contain $\mu = 10$. Again, your results may differ.

(c) According to theory, we would expect 95% of the 60 confidence intervals to contain $\mu = 10$; in other words $60(.95) = 57$ of 60. So, my simulation results are slightly different than what the theory suggests.

---

3. Problem 2.16, p. 77, using $\sigma = 6.02$.

NOTE: It's fine if this question was done using Minitab.

(a) 99% CI for μ, the mean stopping distance.
We are told that $\bar{y} = 57.8$, $\sigma = 6.02$ and $n = 81$.
For a 99% CI, we need $z_{\alpha/2}$, where $\alpha = 1 - .99 = 0.01$, so $z_{\alpha/2} = z_{.01/2} = z_{.005}$.
The quick way to find $z_{0.005}$ using the table in our text is to look up $0.5 - 0.005 = 0.495$ in the **middle** of the normal table (Table A1 in text). The closest value to 0.495 is is either 2.57 or 2.58 (either is acceptable). Therefore, $z_{.005} = 2.58$.
The 99% CI for μ is

$$\bar{y} \pm z_{.005}\frac{\sigma}{\sqrt{n}} = 57.8 \pm 2.58(6.02)/\sqrt{81} = (57.8 - 1.73, 57.8 + 1.73) = (56.07, 59.53)$$

(b) 2.19, p. 77.
NOTE: It's fine if this was done in Minitab.
$H_o : \mu = 60$ vs. $H_o : \mu < 60$
This is a lower-tailed test.

$$\text{Test stat: } z_{obs} = \frac{\bar{y} - \mu_o}{\sigma/\sqrt{n}} = \frac{57.8 - 60}{6.02/\sqrt{81}} = -3.29$$

Test at $\alpha = 0.05$. There are two ways to approach this. First, I'll use the p-value approach (since you are asked to find the p-value, anyway). Since this is a lower-tailed test:

$$\text{p-value } = P(z < -3.29) = P(z > 3.29)$$

since the (standard) normal distribution is symmetric about 0.
Now,

$$P(z > 3.29) = P(z > 0) - P(0 < z < 3.29) = 0.5 - .499 = 0.001(\text{approximately})$$

Therefore p-value = 0.001.
It is approximate since the table doesn't go any higher than 3.29, so we use 3.09 (the table's largest value) to give us the 0.499.
Since $0.001 < \alpha = 0.05$, we reject H_o.
It appears that mean stopping distances are less than 60 feet.
Second approach: Reject H_o if $z_{obs} < -z_\alpha$.
To find $z_\alpha = z_{0.05}$: look up (0.5 - 0.05) = 0.45 in middle of normal table. Then $z_{0.05} = 1.65$ (1.64 also acceptable).
Since $-3.29 < -1.65$, reject H_o.

4. Refer to 3.1, p. 125, and complete by hand.

   (a) 3.1(a): Scatterplot of data.
   3.1(b): The scatterplot suggests there is a strong (positive) linear relationship between GPA (x) and starting salary (y). Therefore, a simple linear regression model would be appropriate to relate y to x.

   (b) Find $\hat{\beta}_0$ and $\hat{\beta}_1$ the hard way: by hand.
   First, some summary stats of the $n = 7$ observations:

$$\sum_i x_i = 3.26 + \ldots + 3.47 = 21.57, \quad \bar{x} = 21.57/7 = 3.08,$$

$$\sum_i y_i = 33.8 + \ldots + 35.3 = 226.8, \quad \bar{y} = 226.8/7 = 32.4$$

$$\sum_i x_i^2 = (3.26)^2 + \ldots + (3.47)^2 = 68.31, \quad \sum_i y_i^2 = (33.8)^2 + \ldots + (35.3)^2 = 7409.7,$$

$$\sum_i x_i y_i = (3.26)(33.8) + \ldots + (3.47)(35.3) = 709.37$$

   Then

$$SS_{xy} = \sum_i x_i y_i - \frac{1}{n}\sum_i x_i \sum_i y_i = 709.37 - \frac{1}{7}(21.57)(226.8) = 10.50$$

$$SS_{xx} = \sum_i x_i^2 - \frac{1}{n}\left(\sum_i x_i\right)^2 = 68.31 - \frac{1}{7}(21.57)^2 = 1.84$$

   Therefore

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{10.50}{1.84} = 5.71, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 32.4 - 5.71(3.08) = 14.81$$

   and the least squares regression line is

$$\hat{y} = 14.81 + 5.71x$$

   (c) Residual if GPA = 2.60.
   From data: if $x_i = 2.60$, $y_i = 29.8$.
   From least squares line: $\hat{y} = 14.81 + 5.71(2.6) = 29.66$.
   Residual: $e_i = y_i - \hat{y}_i = 29.8 - 29.66 = 0.14$.

---

5. 3.5, p. 125–126, using Minitab or the software package of your choice. Also report $\hat{\beta}_1$, R^2, r and $\hat{\sigma}^2$.

   From the plot, there appears to be a strong (positive) linear relationship between the time required for a service call and the number of copiers serviced. Therefore, a simple linear regression would be appropriate to describe the relationship.

   The Minitab output is below:

Minitab Project Report

Regression Analysis: Minutes versus Copiers

The regression equation is
Minutes = 11.5 + 24.6 Copiers

Predictor	Coef	SE Coef	T	P
Constant	11.464	3.439	3.33	0.009
Copiers	24.6022	0.8045	30.58	0.000

S = 4.61521   R-Sq = 99.0%   R-Sq(adj) = 98.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	19919	19919	935.15	0.000
Residual Error	9	192	21		
Total	10	20111			

We see that $\hat{\beta}_1 = 24.60$, $\hat{\sigma}^2 = s^2 = (4.61)^2 = 21.25$, $R^2 = 0.99 = 99\%$ and $r = +\sqrt{0.99} = 0.995$, since the sign of $\hat{\beta}_1$ is positive.