# STATISTICS
## Homework 7

**Due Jan. 10, 2011**


1. A factory has thirty production lines for a specific product. All production lines are run under the same process and are independently operated. A new measure was introduced to reduce the product defect ratio from these production lines. The following data represent the defect ratios of these production lines, before and after the measure was implemented.

| Production Line | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Before | **8.05** | **6.00** | **3.50** | **5.21** | **5.06** | **5.37** | **5.96** | **8.49** | **0.44** | **3.63** |
| After | **6.50** | **2.21** | **5.17** | **4.31** | **1.34** | **3.75** | **4.75** | **8.14** | **0.67** | **1.87** |
| Production Line | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Before | **7.99** | **4.07** | **5.55** | **5.97** | **4.92** | **5.35** | **3.18** | **6.83** | **6.39** | **6.42** |
| After | **5.26** | **3.83** | **2.98** | **1.41** | **5.75** | **5.02** | **3.07** | **3.90** | **4.67** | **5.71** |
| Production Line | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| Before | **4.93** | **6.21** | **5.74** | **4.17** | **2.39** | **1.94** | **6.95** | **7.32** | **6.27** | **6.30** |
| After | **3.90** | **3.71** | **4.17** | **2.88** | **2.97** | **1.54** | **7.44** | **5.06** | **2.98** | **2.84** |

Conduct a test to see whether implementation of the measure does reduce the defect ratio. Use level of significance $\alpha = 0.05$.

(1) What is the null hypothesis?

(2) What is your test statistic?

(3) What is the critical region of the test statistic?

(4) What is the p-value of the test statistic based on the above data?

(5) What is your conclusion based on the result of your test?

2. The following data represents a random sample from an exponential density with parameter $\lambda$ $f_X(x;\lambda) = \lambda e^{-\lambda x} I_{[0,\infty)}(x)$ , $\lambda > 0$.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 9.73 | 5.12 | 7.65 | 4.78 | 5.81 | 0.75 | 5.12 | 2.90 | 0.90 | 5.70 |
| 10.25 | 3.90 | 3.11 | 3.85 | 0.84 | 3.59 | 0.34 | 18.78 | 5.01 | 4.73 |

Conduct a hypothesis test to see whether the null hypothesis $H_0 : \lambda \le 0.1$ ($H_1 : \lambda > 0.1$) should be rejected at level of significance $\alpha$=0.05. [Hint: For an exponential random variable $X$ with parameter $\lambda$, the quantity $2n\lambda\bar{x}_n$ is a pivotal quantity having a chi-squared distribution of degree of freedom $2n$.]

3. The following is the output of a t-test from R.

```
> x=rnorm(          )
> t.test(x,mu=75,alt="less")

        One Sample t-test

data:  x
t =          , df = 39, p-value = 0.02745
alternative hypothesis: true mean is less than 75
95 percent confidence interval:
     -Inf 74.26614
sample estimates:
mean of x
 70.06415
```

(1) What is the null hypothesis?

(2) What is the sample size of the random sample?

(3) What is the value of the test statistic calculated from the random sample?

(4) What is the sample standard deviation ($S_n$) of the random sample?

(5) Will the null hypothesis be rejected at level of significance $\alpha$=0.05?

4. Data in the following table represents a random sample of final grades of students of statistics class over a few years.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 73 | 61 | 51 | 69 | 66 | 67 | 84 | 87 | 77 | 57 |
| 81 | 61 | 30 | 23 | 74 | 67 | 64 | 64 | 49 | 35 |
| 77 | 71 | 70 | 49 | 68 | 52 | 77 | 62 | 62 | 30 |
| 65 | 62 | 66 | 81 | 71 | 73 | 69 | 27 | 83 | 69 |
| 72 | 64 | 71 | 67 | 82 | | | | | |

Suppose that final grades of students in statistics class are normally distributed. At level of significance $\alpha$=0.05, test whether the mean grade is lower than 65 $(H_0 : \mu \geq 65,\ \ H_1 : \mu < 65)$. Develop and draw the power function for the test.

5. The data in the following table represent a set of sample pairs of X and Y.

| | | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **00** | **X** | **1.28** | **1.85** | **3.05** | **8.63** | **249.63** | **480.45** | **486.83** | **462.60** | **392.90** | **440.93** |
| | **Y** | **0.58** | **1.24** | **1.79** | **136.82** | **132.63** | **241.98** | **465.73** | **471.92** | **448.43** | **380.87** |
| **10** | **X** | **427.33** | **362.30** | **316.18** | **290.20** | **357.85** | **413.73** | **415.43** | **432.85** | **418.40** | **441.78** |
| | **Y** | **427.42** | **414.24** | **351.20** | **306.50** | **281.31** | **346.89** | **401.06** | **402.71** | **419.59** | **405.58** |
| **20** | **X** | **784.47** | **951.78** | **735.03** | **647.05** | **638.98** | **525.93** | **401.83** | **292.18** | **216.20** | **169.78** |
| | **Y** | **750.36** | **760.44** | **922.63** | **712.52** | **627.23** | **619.41** | **509.82** | **389.52** | **283.23** | **209.58** |
| **30** | **X** | **137.63** | **116.95** | **103.80** | **100.20** | **88.30** | | | | | |
| | **Y** | **164.58** | **133.41** | **113.37** | **100.62** | **97.13** | | | | | |

(1) Construct the scatter plot of (x,y).

(2) Conduct a test to see whether the population means of X and Y are different at level of significance $\alpha$ = 0.05.

(3) Plot the time series data of X and Y, respectively.

(4) Plot the time series data of X – Y.

(5) Make your comment on the results of the test in (2).

6. Suppose that tag numbers (such as TP-000266, TP-018302, TP-010135, TP-020362, …) of cars registered in a city are issued in series (starting from, for example, TP-000000). A market investigator was asked to estimate the total number of cars registered in the city. The investigator is a person with good knowledge of statistics. He went to the parking lot of a popular shopping mall and recorded the tag numbers of cars (see the following table). If you were the market investigator,
   (1) What is your estimate of the 95% confidence interval of the number of cars registered in the city?
   (2) Test whether the total number of cars registered in the city is less than 200000.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 122693 | 7594 | 187456 | 5060 | 197346 | 61076 | 140399 | 194731 | 85065 | 182 |
| 187119 | 135746 | 45155 | 89863 | 22326 | 65830 | 69799 | 105739 | 16053 | 133880 |
| 50362 | 2385 | 197837 | 69965 | 21549 | 157963 | 26206 | 25125 | 111711 | 134691 |
| 134553 | 32692 | 5175 | 142525 | 132373 | 94143 | 58099 | 43366 | 24127 | 157714 |
| 39273 | 119823 | 180481 | 119286 | 115727 | 87632 | 122877 | 178230 | 23243 | 17852 |
| 97689 | 190281 | 152152 | 35837 | 38336 | 124707 | 42108 | 132952 | 1635 | 65394 |
| 124429 | 169923 | 104172 | 159055 | 29018 | 171051 | 29529 | 127162 | 34833 | 75546 |
| 113066 | 90822 | 169718 | 150785 | 82014 | 56582 | 92386 | 46322 | 146039 | 185257 |
| 46513 | 136116 | 137508 | 133769 | 129354 | 173511 | 158352 | 112715 | 149055 | 157381 |
| 161837 | 9665 | 114446 | 143571 | 71506 | 76237 | 152296 | 102906 | 13794 | 41818 |

7. Final grades of students in a statistics class are listed in the following table.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 69 | 46 | 51 | 34 | 85 | 79 | 49 | 58 | 52 | 57 | 50 |
| 39 | 74 | 46 | 57 | 45 | 49 | 44 | 60 | 53 | 59 | 24 |
| 75 | 30 | 70 | 44 | 20 | 34 | 84 | 45 | 55 | 57 | 53 |
| 54 | 35 | 38 | 49 | 52 | 45 | 79 | 68 | 63 | 57 | 65 |

   (1) At level of significance $\alpha = 0.05$, test whether the grade distribution is significantly different from a normal distribution with mean $\mu = 60$ and standard deviation $\sigma = 15$, using the ***chisq.test*** function in R. Set the number of categories equal to 6.
   (2) An exemplar R code for chi-square GOF test using equiprobable categories can be downloaded from the class webpage. Modify the exemplar code and use it to test, at level of significance $\alpha = 0.01$, whether the final grades are not exponentially distributed. Set the number of categories equal to 6.

8. The Kolmogorov-Smirnov GOF test can be implemented using ***ks.test*** in R. The ks.test is rather straightforward for one-sample case. A brief description is as follows:

   *ks.test(x, y, parameters, alternative="two.sided")*

   where *x* is the data vector to be tested, *y* is a string vector specifying the hypothesized distribution, *parameters* are the values of distribution parameters corresponding to *y*, and *alternative* represents a string vector ("*less*", "*greater*", or "*two.sided*") for one-tail or two-tail test.

   Examples:

   *ks.test(x, "pnorm", 30, 10, alternative="two.sided")*
   *ks.test(x, "pexp", 0.2, alternative="greater")*

At level of significance $\alpha = 0.05$, test whether data in the following table is significantly different from a normal distribution with mean 60 and standard deviation 15.

| | | | | |
|---|---|---|---|---|
| 51.29 | 53.45 | 75.82 | 49.08 | 81.47 |
| 49.38 | 58.61 | 66.07 | 65.27 | 55.49 |
| 43.85 | 44.04 | 57.11 | 63.92 | 80.54 |
| 42.82 | 72.91 | 78.42 | 71.39 | 81.40 |