Math 207
Introduction to Statistics

Final Exam

No books or notes may be used. You may, however, use a calculator or statistical software (such as R).

1. (10 points) A program is being tested on an extremely expensive computer. Due to costs, the program is only run twice. The output values are 4 and 8. If the program is running correctly, its average output value is 4. Is the program running correctly? State the statistical test that you use and compute the test statistic. Estimate the P-value and discuss its significance.

2. (10 points) Suppose 25 readings on the length of a bike trail show an average of 12 miles and SD of 50 feet.

a) The average of all 25 measurements is likely to be off the exact length by ______________ or so.

b) Esimate the probability that the average of all 25 measurements will be within 10 feet of the exact length.

c) Find a 95% confidence interval on the length of the trail.

3. (20 points) A simple random sample of 144 Pittsburgh residents is taken. 1/2 of these residents indicate that they own hybrid-electric vehicles.

a) Find a 95% confidence interval on the percentage of residents who own hybrid-electric vehicles.

b) Explain what this 95% confidence interval means (i.e., there is a 95% chance of what?).

c) Do the data support the null hypothesis that 45% of the Pittsburgh residents own hybrid-electric vehicles? (i.e., that the observed value differs from 45% due to chance)? State the statistical test that you use and estimate the P-value.

d) Suppose you learn that all 144 of the subjects were medical administrators. How might this impact your answers to b) and c)?

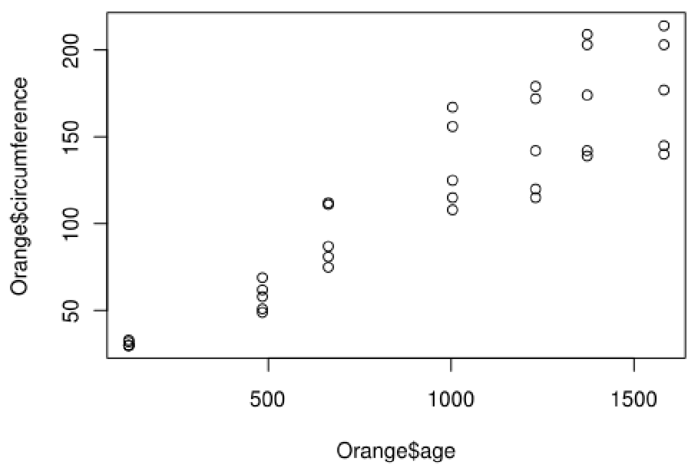
4. (10 points) A simple random sample of 64 Winchester residents is taken. 1/5 of these residents indicate that they own hybrid-electric vehicles. Use this data and the data from #3 to test the hypothesis that the proportions of Winchester and Pittsburgh residents who own hybrid-electric vehicles are the same. Estimate the P-value and discuss its significance.

5. (10 points) A coin is to be flipped 13 times. What is the chance of getting exactly 11 heads? Simplify your answer as much as possible. What is the chance of getting at least 11 heads?

6. (15 points) Data is collected for the age (in days) and circumference (in millimeters) of orange trees. The following statistics summarize the data

$$\text{average age} = 900 \text{ days} \qquad \text{SD}_{\text{days}} = 500 \text{ days}$$

$$\text{average circumference} = 120 \text{ mm} \qquad \text{SD}_{\text{circ}} = 60 \text{ mm} \qquad r = 0.7$$



a) Find a formula for the regression line to predict circumference from age. Sketch the line in the figure above.

b) Estimate the circumference of an orange tree that is 1035 days old.

c) About 68% of the orange trees that are 1035 days old have what range of circumferences?

7. (5 points) Suppose that you take a large number of independent random samples from a box of numbered tickets. Fill in the blanks.

a) The histogram of the averages of the samples will will look approximately like a __________________.

b) The histogram of the sums of the samples will will look approximately like a __________________.

c) Suppose that the tickes are all 0s and 1s. The histogram of the percent of 1s in the samples will will look approximately like a __________________.

8. (5 points) 1. A fair die is rolled 10 times. The number of dots showing on the rolls was: 2, 1, 4, 4, 2, 4, 1, 3, 1, 5. Draw a histogram of this data.

9. (15 points) An experiment consists of rolling a die 400 times and recording how often the die shows an even number.

a) Write down a box model to represent the experiment.

b) Compute the mean and SD of the box.

c) The number of times that the die will show an even number will be around ________________,

give or take ________________ or so.

d) The chance that the percent of times will be between 45% and 55% is about ________________ %.

Formula Sheet

$$\text{mean} = \frac{1}{n}\left(x_1 + \cdots + x_n\right)$$

$$\text{SD} = \sqrt{\frac{1}{n}\left((x_1 - \text{mean})^2 + \cdots + (x_n - \text{mean})^2\right)}$$

$$\text{sd} = \sqrt{\frac{1}{n-1}\left((x_1 - \text{mean})^2 + \cdots + (x_n - \text{mean})^2\right)}$$

$$\text{EV}_{\text{sum}} = n\,\text{AV}_{\text{box}}$$
$$\text{SE}_{\text{sum}} = \sqrt{n}\;\;\text{SD}_{\text{box}}$$

$$\text{EV}_{\text{av}} = \text{AV}_{\text{box}}$$
$$\text{SE}_{\text{av}} = \text{SD}_{\text{box}} \big/ \sqrt{n}$$

Shortcut formula (if there are only two different kinds of numbers in the box):

$$\text{SD}_{\text{box}} = (\text{big }\# - \text{small }\#)\sqrt{\begin{pmatrix}\text{fraction of}\\ \text{tickets with the}\\ \text{big number}\end{pmatrix}\begin{pmatrix}\text{fraction of}\\ \text{tickets with the}\\ \text{small number}\end{pmatrix}}$$

$$95\%\text{ confidence interval} = \text{observed } \pm\ 2\,\text{SE}$$

$$z = \frac{\text{observed} - \text{expected}}{\text{standard error}}$$

Regression line:

$$y - y_{\text{av}} = r\,\frac{\text{sd}_y}{\text{sd}_x}\;(x - x_{\text{av}})$$

RMS error for the regression line:

$$\text{RMS}_{\text{reg}} = \text{sd}_y\,\sqrt{1 - r^2}$$

Binomial Formula:

$$\frac{n!}{k!\,(n-k)!}\,, p^k\,(1-p)^{n-k}$$

Math 207
Introduction to Statistics

Final Exam

No books or notes may be used. You may, however, use a calculator or statistical software (such as R).

1. (10 points) A program is being tested on an extremely expensive computer. Due to costs, the program is only run twice. The output values are 1 and 5. If the program is running correctly, its average output value is 1. Is the program running correctly? State the statistical test that you use and compute the test statistic. Estimate the P-value and discuss its significance.

2. (10 points) Suppose 16 readings on the length of a bike trail show an average of 8 miles and SD of 60 feet.

a) The average of all 16 measurements is likely to be off the exact length by ______________ or so.

b) Esimate the probability that the average of all 16 measurements will be within 30 feet of the exact length.

c) Find a 95% confidence interval on the length of the trail.

3. (20 points) A simple random sample of 225 Pittsburgh residents is taken. 1/2 of these residents indicate that they own hybrid-electric vehicles.

a) Find a 95% confidence interval on the percentage of residents who own hybrid-electric vehicles.

b) Explain what this 95% confidence interval means (i.e., there is a 95% chance of what?).

c) Do the data support the null hypothesis that 45% of the Pittsburgh residents own hybrid-electric vehicles? (i.e., that the observed value differs from 45% due to chance)? State the statistical test that you use and estimate the P-value.

d) Suppose you learn that all 225 of the subjects were medical administrators. How might this impact your answers to b) and c)?

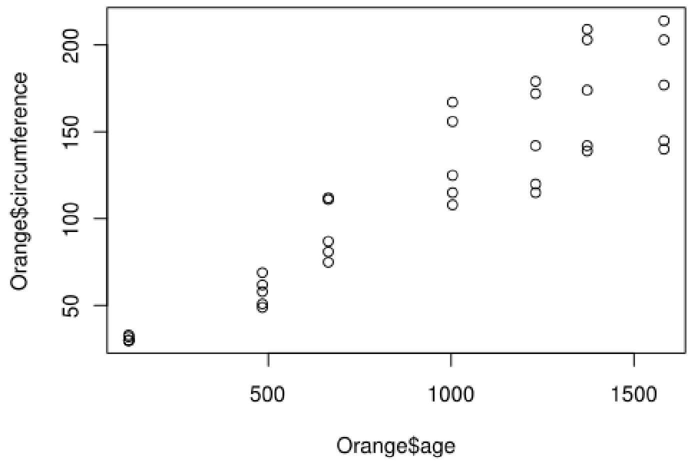
4. (10 points) A simple random sample of 144 Winchester residents is taken. 2/5 of these residents indicate that they own hybrid-electric vehicles. Use this data and the data from #3 to test the hypothesis that the proportions of Winchester and Pittsburgh residents who own hybrid-electric vehicles are the same. Estimate the P-value and discuss its significance.

5. (10 points) A coin is to be flipped 12 times. What is the chance of getting exactly 11 heads? Simplify your answer as much as possible. What is the chance of getting at least 11 heads?

6. (15 points) Data is collected for the age (in days) and circumference (in millimeters) of orange trees. The following statistics summarize the data

$$\text{average age} = 900 \text{ days} \qquad \mathrm{SD}_{\text{days}} = 500 \text{ days}$$

$$\text{average circumference} = 120 \text{ mm} \qquad \mathrm{SD}_{\text{circ}} = 60 \text{ mm} \qquad r = 0.8$$



a) Find a formula for the regression line to predict circumference from age. Sketch the line in the figure above.

b) Estimate the circumference of an orange tree that is 1000 days old.

c) About 68% of the orange trees that are 1000 days old have what range of circumferences?

7. (5 points) Suppose that you take a large number of independent random samples from a box of numbered tickets. Fill in the blanks.

a) The histogram of the averages of the samples will will look approximately like a ________________.

b) The histogram of the sums of the samples will will look approximately like a ________________.

c) Suppose that the tickes are all 0s and 1s. The histogram of the percent of 1s in the samples will will look approximately like a ________________.

8. (5 points) 1. A fair die is rolled 10 times. The number of dots showing on the rolls was: 2, 1, 3, 4, 2, 4, 1, 3, 6, 5. Draw a histogram of this data.

9. (15 points) An experiment consists of rolling a die 144 times and recording how often the die shows an even number.

a) Write down a box model to represent the experiment.

b) Compute the mean and SD of the box.

c) The number of times that the die will show an even number will be around ______________,

give or take ______________ or so.

d) The chance that the percent of times will be between 45% and 55% is about ______________ %.

Formula Sheet

$$\text{mean} = \frac{1}{n}\,(x_1 + \cdots + x_n)$$

$$\text{SD} = \sqrt{\frac{1}{n}\,((x_1 - \text{mean})^2 + \cdots + (x_n - \text{mean})^2)}$$

$$\text{sd} = \sqrt{\frac{1}{n-1}\,((x_1 - \text{mean})^2 + \cdots + (x_n - \text{mean})^2)}$$

$$\text{EV}_{\text{sum}} = n\,\text{AV}_{\text{box}}$$
$$\text{SE}_{\text{sum}} = \sqrt{n}\ \ \text{SD}_{\text{box}}$$

$$\text{EV}_{\text{av}} = \text{AV}_{\text{box}}$$
$$\text{SE}_{\text{av}} = \text{SD}_{\text{box}} \big/ \sqrt{n}$$

Shortcut formula (if there are only two different kinds of numbers in the box):

$$\text{SD}_{\text{box}} = (\text{big }\# - \text{small }\#)\ \sqrt{\left(\begin{array}{c}\text{fraction of}\\ \text{tickets with the}\\ \text{big number}\end{array}\right)\left(\begin{array}{c}\text{fraction of}\\ \text{tickets with the}\\ \text{small number}\end{array}\right)}$$

$$95\%\text{ confidence interval} = \text{observed}\ \pm\ 2\,\text{SE}$$

$$z = \frac{\text{observed} - \text{expected}}{\text{standard error}}$$

Regression line:

$$y - y_{\text{av}} = r\,\frac{\text{sd}_y}{\text{sd}_x}\ (x - x_{\text{av}})$$

RMS error for the regression line:

$$\text{RMS}_{\text{reg}} = \text{sd}_y\,\sqrt{1 - r^2}$$

Binomial Formula:

$$\frac{n!}{k!\,(n-k)!}\,, p^k\,(1-p)^{n-k}$$