

Final Exam

STAT 105, Section B FALL 2015

Instructions

- The exam is scheduled for 120 minutes, from 7:30 to 9:30 AM. At 9:30 AM the exam will end.
- A formula sheet is attached to the end of the exam. Feel free to tear it off.
- You may use a calculator during this exam.
- Answer the questions in the space provided. If you run out of room, continue on the back of the page or on scratch paper.
- If you have any questions about, or need clarification on the meaning of an item on this exam, please ask your instructor. No other form of external help is permitted attempting to receive help or provide help to others will be considered cheating.
- **Do not cheat on this exam.** Academic integrity demands an honest and fair testing environment. Cheating will not be tolerated and will result in an immediate score of 0 on the exam and an incident report will be submitted to the dean's office.

Name: _____

Student ID: _____

1. (2 points) (TRUE/FALSE) A random sample of 1000 student's Statistics exam scores was drawn from the population of all possible Stat scores (an unknown distribution). Once the sample mean is computed, it can be viewed as the distribution/population mean.
2. (2 points) (TRUE/FALSE) While trying to figure out the probability that the sample mean for a data of size 10 would exceed a value, we can apply the central limit theorem.
3. A sample of size 3 was drawn from a population and the resulting observations are reported below.

2.2, 2, 10.3

Using these observed values, report the following:

- (a) (2 points) the mean
- (b) (2 points) the median
- (c) (2 points) the variance
- (d) (2 points) the standard deviation

4. An agriculturist is attempting to determine which of three species of corn (A, B, and C) yield the most grain per acre. Since the yield may depend on the fertilizer used, the researcher intends to use fertilizers with different concentrations of Nitrogen as well - low Nitrogen, medium-low Nitrogen, medium-high Nitrogen, and high Nitrogen. There are 8 fields (scattered around Iowa) available to perform this experiment. Each field is divided into 24 single acre plots and the combinations of species and fertilizer are randomly assigned so that within each field every combination is used exactly twice. At harvest time, the amount of grain each plot yields is recorded and the combination of corn species and fertilizer that gives the highest average yield is chosen.

(a) (2 points) Is this an experiment or an observational study? Explain.

(b) Identify the following (if there was not one, simply put "not used"). Additionally, label each as continuous or discrete.

i. (2 points) Response variable(s):

ii. (2 points) Experimental variable(s):

iii. (2 points) Blocking variable(s):

(c) (2 points) Was replication used in this study? If so, where was it applied? If not, how could we have applied it?

5. A company specializing in the installation and maintenance of “infinity pools” records the number of service requests they receive each month for two years. The number of requests are presented in the table below:

The decimal point is 1 digit(s) to the right of the |

0 | 122333444

0 | 556669999

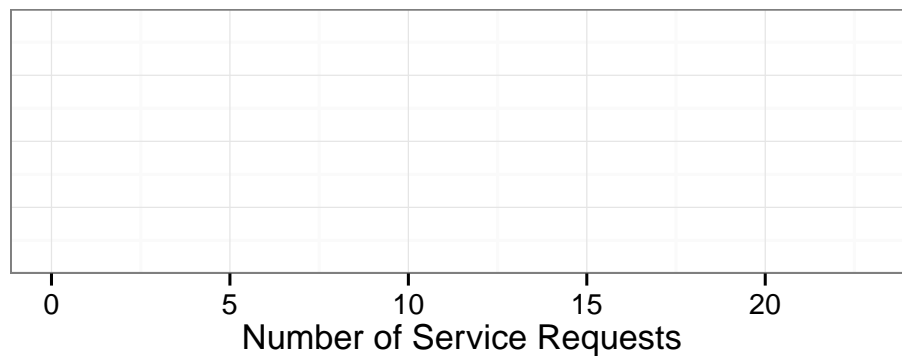
1 | 4

1 | 77

2 | 013

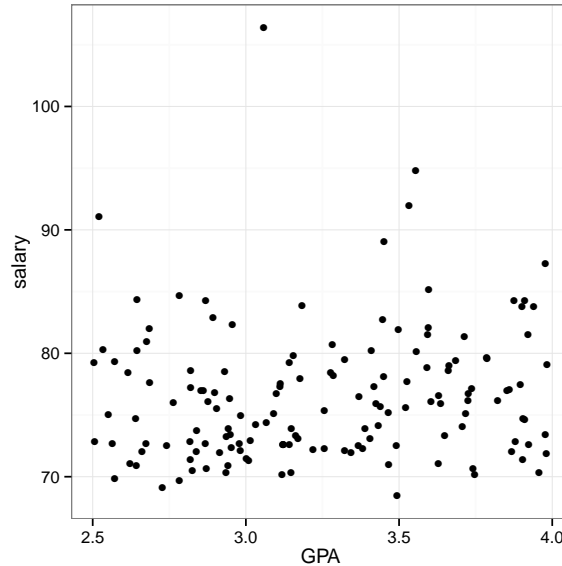
Note that 0 | 4 represents 4 and 1 | 2 represents 12. In this case, the first quartile is $Q(.25) = 3.5$, the median is 6, and the third quartile is $Q(.75) = 11.5$.

- (a) (10 points) Using the axes below, create a box plot to summarize the data. Label all important values. Draw a star over unusual observations.



6. A survey given to members of a national engineering honor society who have recently graduated is attempting to determine the relationship between salary and GPA. The graph below displays 150 responses.

The results are depicted below (using GPA on the x-axis):



Here are some summaries of the data (again using the actual score as the x-value and the person's evaluation of their score as the y-value):

$$\sum_{i=1}^{150} x_i = 487$$

$$\sum_{i=1}^{150} x_i^2 = 1609$$

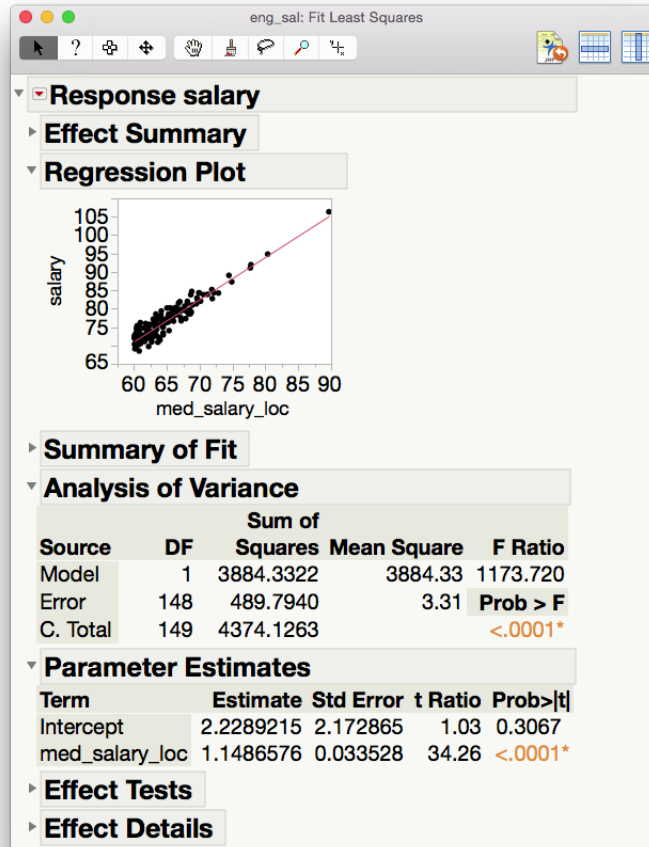
$$\sum_{i=1}^{150} y_i = 11474$$

$$\sum_{i=1}^{150} y_i^2 = 882126$$

$$\sum_{i=1}^{150} x_i y_i = 37299$$

- (a) Using the summaries above, the survey workers fit a linear relationship between **GPA** (x) and **salary** (y).
- (5 points) Write the equation of the fitted linear relationship.
 - (5 points) Using the fitted line, what do we suppose the salary will be for an engineer with a GPA of 3.0?

- (b) Discouraged by the relationship between salary and GPA, the surveyors remember that they know the address of each respondent and are able to determine the median income of the area in which the respondent lives. The JMP output below comes from fitting a linear relationship using the annual salary of the respondent (“salary”) and the median income of the area in which the respondent lives (med_salary_loc).



- i. (5 points) Write the equation of the fitted linear relationship.
- ii. (5 points) Find and interpret the value of R^2 for the fitted quadratic relationship.
7. A winery is experimenting with blending a small amount of non-grape tastes into its current harvest of grapes (to add “notes”). They are considering three fruit additions (apple, cherry, and kiwi) and two spice additions (oak and vanilla). Three wine experts working for the company test the fruit/spice combinations and provide a rating from 0 to 10 (with 10 being the highest). The results are recorded below.

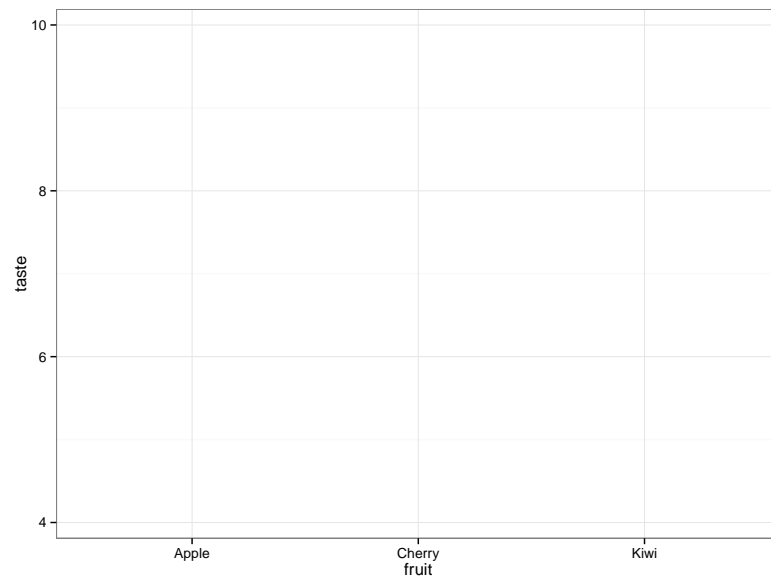
Fruit	Spice	
	Oak	Vanilla
Apple	9.8	8.4
	9.9	8.9
	9.3	8.3
Cherry	8.3	5.3
	5.5	4.9
	5.7	4.7
Kiwi	8	7
	8.1	7.3
	7.6	6.6

The following summaries may help in this problem:

Fruit	Spice		
	Oak	Vanilla	
Apple	$\bar{y}_{11} = 9.67$	$\bar{y}_{12} = 8.53$	$\bar{y}_{1\cdot} = 9.1$
Cherry		$\bar{y}_{2\cdot} = 5.04$	
Kiwi	$\bar{y}_{31} = 7.9$	$\bar{y}_{32} = 6.97$	$\bar{y}_{3\cdot} = 7.44$
	$\bar{y}_{\cdot 1} = 7.69$		$\bar{y}_{\cdot\cdot} = 7.19$

- (a) (2 points) Report the value of \bar{y}_{21}
- (b) (2 points) Report the value of $\bar{y}_{\cdot 2}$
- (c) (3 points) Find the fitted main effect of fruit, a_1 , a_2 , and a_3 , that you would get from factorial model that ignores interactions.
- (d) (3 points) Ignoring possible interactions, give the estimated values \hat{y}_{22} and \hat{y}_{23} .
- (e) (2 points) How do the estimated values computed above compare to the average for the same combinations seen in the data? Does it appear that ignoring interactions was a good choice?

(f) (5 points) Using the template below, create a profile plot for this data:



(g) (2 points) Using the plot does it appear that there are interactions between fruit and spice type?
Which combination would you recommend?

8. Let X be a normal random variable with a mean of 1 and a variance of 16 (i.e., $X \sim N(1, 16)$) and let Z be a random variable following a standard normal distribution. Find the following probabilities (note: the attached standard normal probability table may be helpful):

(a) (2 points) $P(Z \leq 1)$

(b) (2 points) $P(|Z| \geq 2)$

(c) (2 points) $P(0 \leq X < 5)$

(d) (2 points) $P(|X| \leq 5)$

9. Suppose that X is a continuous random variable with cumulative density function (cdf):

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-3x} & x \geq 0 \end{cases}$$

(a) (2 points) What is the probability that X takes a value less than 1?

(b) (2 points) What is the probability that X takes a value greater than 2?

(c) (2 points) Derive $f(x)$, the probability density function.

10. Suppose we have a bag containing five tiles, three of which are labelled 1 and two of which are labelled 2. Assume that each of the five tiles has an equal chance of being drawn. The number on the tile tells us how many times we will roll a fair die. For instance, if we draw a tile with the number 2 on it, we will roll a die twice but if we draw a tile with a 1 on it, we will only roll a die once. For this problem,

- let X be the number on the tile
- let Y be the sum of our rolls

(a) (2 points) Find $f_X(x)$.

(b) (2 points) Find $f_{Y|X}(6|1)$.

(c) (2 points) Find the joint probability $f(1, 6)$.

(d) (2 points) Find the joint probability $f(2, 6)$.

(e) (3 points) Find $f_Y(6)$.

(f) (2 points) Find $f_{X|Y}(2|6)$.

11. O-rings are elastomer loops designed to create a seal between the interface of two parts of a mechanical device. Because the elasticity of the material used to make them can be impacted by temperature (which can lead to the seal being broken) it is important to make sure that the O-ring is functional at the temperatures the part they are used in will be exposed to. Two composites (Composite X and Composite Y) are being tested in an O-ring that will be used in a part of a satellite that will be exposed to very low temperatures. A sample of 50 O-rings from each composite are placed in a chamber, where the temperature is gradually reduced until the seal is broken. Suppose that each composite has some mean failure temperature, μ_X for Composite X and μ_Y for Composite Y, and some variance in failure temperature, σ_X^2 for Composite X and σ_Y^2 for Composite Y. Before any observations are recorded, we can consider the sampled values from Composite X to be random variables X_1, X_2, \dots, X_{50} with $\mathbb{E}(X_i) = \mu_X$ and $\text{Var}(X_i) = \sigma_X^2$. We can also consider the sampled values from Composite Y to be random variables Y_1, Y_2, \dots, Y_{50} with $\mathbb{E}(Y_i) = \mu_Y$ and $\text{Var}(Y_i) = \sigma_Y^2$.

Let $\bar{X} = \frac{1}{50}X_1 + \frac{1}{50}X_2 + \dots + \frac{1}{50}X_{50}$ and let $\bar{Y} = \frac{1}{50}Y_1 + \frac{1}{50}Y_2 + \dots + \frac{1}{50}Y_{50}$.

(a) (3 points) What is the expected value of \bar{X} (use appropriate symbols if needed).

(b) (3 points) What is the variance of \bar{X} (use appropriate symbols if needed).

(c) (3 points) What is the distribution of \bar{X} (use appropriate symbols if needed).

(d) (3 points) What is the expected value of \bar{Y} (use appropriate symbols if needed).

(e) (3 points) What is the variance of \bar{Y} (use appropriate symbols if needed).

(f) (3 points) What is the distribution of \bar{Y} (use appropriate symbols if needed).

(g) (6 points) Let $\bar{D} = \bar{X} - \bar{Y}$. What is the distribution of \bar{D} (use appropriate symbols if needed).

12. After running the O-ring experiment, the researchers found $\bar{x} = 50$ K and $\bar{y} = 53$ K. Suppose that $\sigma_X^2 = 10$ and $\sigma_Y = 20$.

(a) (4 points) Provide a 90% confidence interval for μ_X .

(b) (4 points) Provide a 99% confidence interval for μ_X .

(c) (4 points) Provide a 95% confidence interval for μ_Y .

(d) (6 points) Provide a 95% confidence interval for $\mu_X - \mu_Y$ (hint: you can use the distribution of \bar{D}). Does this provide any evidence that one O-ring is better than the other?

(e) (2 points) Is there any evidence that one O-ring is better than the other?

13. A company recently did a major overhaul to their server system hardware and is checking to make sure that there have been no changes in the download speed. The previous download speed had an average of 63.4 Mbps. A systems analyst took 10 readings on the download speeds during the course of a day to check. Her results are below (in Mbps):

63.63, 63.4, 63.51, 63.14, 63.38, 63.35, 63.53, 63.37, 63.53, 63.71

The sample average is 63.45 and the sample variance is 0.026.

- (a) (5 points) Provide a 90% confidence interval for the mean download speed.
- (b) (5 points) Provide a 95% lower confidence bound for the mean download speed.
- (c) (10 points) Conduct a hypothesis test at the 95% confidence level for the null hypothesis $\mu = 63.4$ against the alternative $\mu \neq 63.4$. Include your hypothesis statement, the test statistic, the p-value, your decision rule, and your conclusion.