Exam I

STAT 105, Section B
FALL 2015

Instructions

- The exam is scheduled for 80 minutes, from 8:00 to 9:20 AM. At 9:20 AM the exam will end.

- A forumula 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.

- 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. Two simple tools are used to measure the speed of an object (in meters per second). By design and after careful callibration, the true speed of the object is *known* to be 10.4 km/s. The measurements of the two tools are below:

   - Measurements from Tool 1: $11.4, 9.3, 8, 8.1, 12.7$

   - Measurements from Tool 2: $7.4, 7.4, 7.5, 7.5, 7.4$

   (a) (2 points) Which tool is better described as accurate?

   (b) (2 points) Which tool is better described as precise?

2. A sample of size 6 was drawn from a population and the resulting observations are reported below.

$$23, 27, 25, 19, 18$$

Using these observed values, report the following:

   (a) (2 points) the mean

   (b) (2 points) the variance

   (c) (2 points) the standard deviation

   (d) (2 points) the value of $Q(.25)$

3. Since it is reasonably quick and cost effecient, the complex paints used to line roads, like many liquids, are routinely shipped by train in tanker cars. However, once the cars arrive at their destination, they often sit for long periods of time before being released to the client - in the case of paint, this waiting time can cause the different components to separate into layers and the paint must be "remixed" before it can be used. A certain city receives shipments of these paints where they are held at the shipping station for a day before being released. A civil engineer working for the city plans to test the effectiveness of 5 methods of remixing the paint. However, there is a small issue: since the tanks themselves are coming from different destinations and through different routes, it is suspected that remixing effectiveness will vary from tank to tank.

Four tanks of paint are currently being shipped. The engineer devises the following plan:

- The content of each tank will be divided into 10 parts.

- Each part will be randomly assigned an ID from 1 - 10.

- Method 1 will be applied to ID 1 and 2, Method 2 will be applied to ID 3 and 4, Method 3 will be applied to ID 5 and 6, Method 4 will be applied to ID 7 and 8, Method 5 will be applied to ID 9 and 10.

The city plans to use the method that most effectively separates the paint in the future.

(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").

    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 experiment? If so, where was it applied? If not, how could we have applied it?

4. New Orleans recently made training and use of body cameras mandatory for police officers. In an **attempt to determine the effectiveness of the change in policy at reducing the number of accusations of police brutality**, a journalist gathered the a week's worth of arrest reports both two months before the policy went into effect and two months after the policy went into effect. For each arrest report, she made note of whether or not the officer was using a body camera, the time of day (measured in hours so that 1:30 pm is 13.5), the difficulty reported in the arrest ("suspect resisted with violence", "suspect resisted without violence", or "suspect did not resist"), and whether or not the suspect made an accusation of brutality after the arrest.

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

(b) (2 points) Identify the response variable(s).

(c) For each of the following variables,

- Identify whether it is qualitative or quantitative variable, and
- If it is qualitative, what are the possible values it can take? If it is quantitative, is it continuous or discrete?

   i. whether or not the office was using a body camera,

   ii. the time of day

   iii. The amount of resistance to arrest

   iv. whether or not the suspect made an accusation of brutality after the arrest

5. A certain make of load bearing beam used in construction of large buildings is certified to withstand pressures of 2.5 tonnes per square inch. 30 beams were tested and the pressures at which they failed are collected in the table below.

```
The decimal point is at the |

1 | 4
2 | 9
3 |
4 | 368899999
5 | 00001111222333444
6 | 45
```

Note that 0 | 9 represents 0.9. In this case, the first quartile is $Q(.25) = 4.9$, the median is 5.05, and the third quartile is $Q(.75) = 5.3$.

(a) (10 points) Complete the following frequency table:

Value Range	Frequency	Relative Frequency	Cumulative Relative Frequency
0.00 - 2.00			
2.01 - 4.00			
4.01 - 6.00			
6.01 - 8.00			

(b) (10 points) Using the axes below, create a box plot to summarize the data. Carefully label the axes.

Pressure at Failure
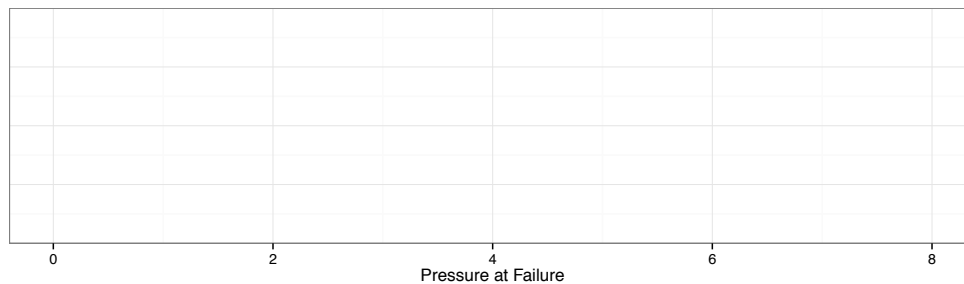
(c) (4 points) Are there any unusually low or high observations? If so, what pressures caused those beams to fail?

(d) (10 points) The company also measured the heat at which the beams begin to deform (in 10,000 degrees Celsius). The values are collected below:

$$28.7, 31.5, 49.4, 44.6, 40.4, 35.1$$

Create a theoretical Q-Q plot using the following quantiles from the normal distribution as the theoretical quantiles. What does this graph tell us about the temperature at which the beams deform?

|	1	2	3	4	5	6
p	0.08	0.25	0.42	0.58	0.75	0.92
$Q(p)$	-1.41	-0.67	-0.2	0.2	0.67	1.41
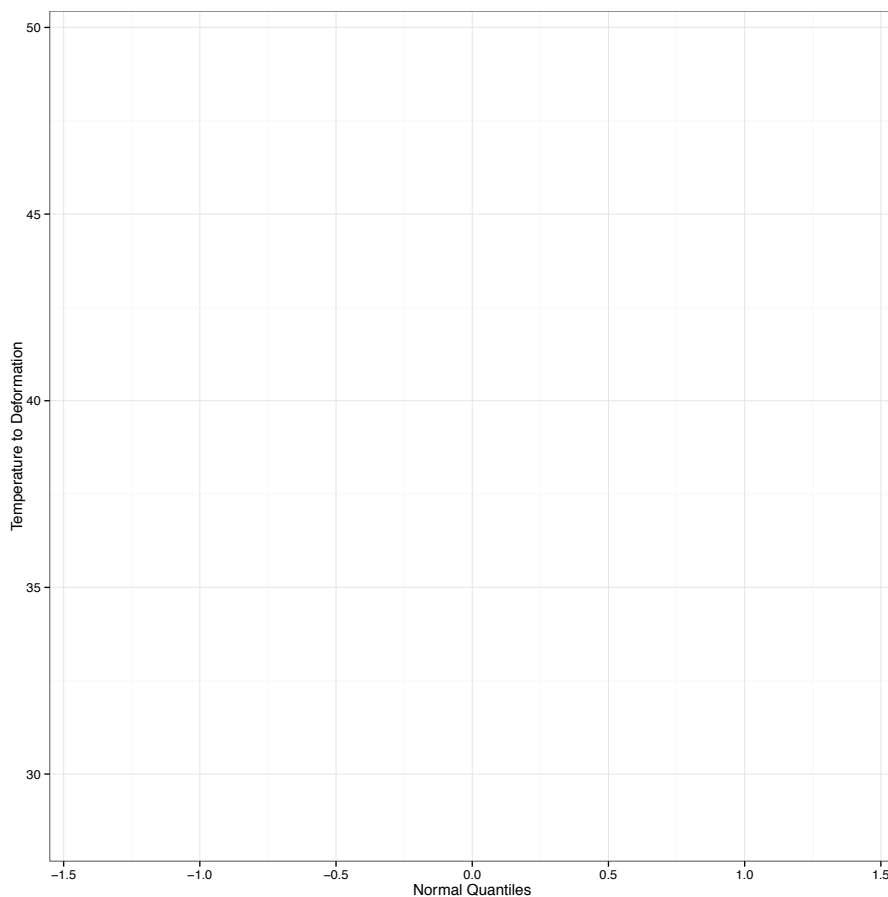
6. Professional engineers often encounter issues relating to *human resources* as they advance in their careers (building a better team of employees is after all not too different than improving any other system, at least on paper). However, many of the "laws" governing human behavior are very different than the strict laws of physics. For instance, a phenomenon known as the Dunning-Kruger effect states that for a given skill incompetent people will

- fail to recognize their own lack of skill
- fail to recognize genuine skill in others
- fail to recognize the extremity of their inadequacy
- recognize and acknowledge their own lack of skill, after they are exposed to training for that skill

A group of 50 job applicants are asked to estimate their skill in technical writing. They are told they will be taking a test with a mean score of 50 and asked to guess what their score will be. Then they are given the test and get an actual score.

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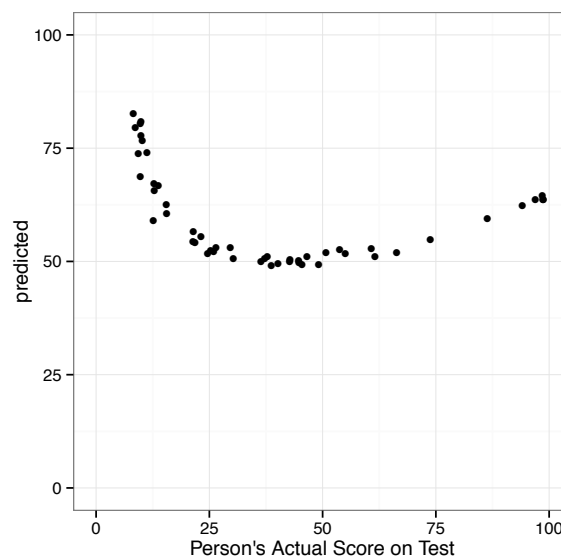
Here are some summaries of the data (again using the actual score as the x-value and the person's evaluation of their score as the y-value):

$$\sum_{i=1}^{50} x_i = 1922 \qquad\qquad \sum_{i=1}^{50} x_i^2 = 110659$$

$$\sum_{i=1}^{50} y_i = 2954 \qquad\qquad \sum_{i=1}^{50} y_i^2 = 179606$$

$$\sum_{i=1}^{50} x_i y_i = 108893$$

(a) Using the summaries above, fit a linear relationship between **the actual score** (x) and **the guessed score** (y).

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 linear relationship.

iii. (5 points) Using the fitted line, what do we suppose a person will guess their score will be if they actually scored a 40.14.

iv. (2 points) A person who scored a 40.14 on the test predicted that they would score 49.56. What is the residual for this person using the linear relationship?

(b) The JMP output below comes from fitting a quadratic model using the actual score (”`actual_score`”) and the square of the actual score (`actual_score^2`).

Response guess_score

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	4031.5440	2015.77	93.5569
Error	47	1012.6597	21.55	**Prob > F**
C. Total	49	5044.2038		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	82.626343	1.864461	44.32	<.0001*
actual_score	-1.246181	0.091868	-13.56	<.0001*
actual_score^2	0.011013	0.000873	12.62	<.0001*

i. (5 points) Write the equation of the fitted quadratic relationship.

ii. (5 points) Find and interpret the value of R^2 for the fitted quadratic relationship.

iii. (5 points) Using the fitted quadratic relationship, what do we suppose a person will guess their score will be if they actually scored a 98.74.

iv. (2 points) A person who scored a 98.74 on the test predicted that they would score 63.55. What is the residual for this person using the quadratic relationship?

STAT 105 Exam I
Reference Sheet

Numeric Summaries

mean $\quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

population variance $\quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$

population standard deviation $\quad \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$

sample variance $\quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$

sample standard deviation $\quad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$

Linear Relationships

Form $\quad y \approx \beta_0 + \beta_1 x$

Fitted linear relationship $\quad \hat{y} = b_0 + b_1 x$

Least squares estimates $\quad b_1 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$

$\quad b_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$

$\quad b_0 = \bar{y} - b_1 \bar{x}$

Residuals $\quad e_i = y_i - \hat{y}_i$

sample correlation coeffecient $\quad r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}}$

$\quad r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)}}$

coeffecient of determination $\quad R^2 = (r)^2$

$\quad \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 - \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$

Multivariate Relationships

Form $\quad y \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$

Fitted relationship $\quad \hat{y} \approx b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$

Residuals $\quad e_i = y_i - \hat{y}_i$

Sums of Squares $\quad \text{SSTO} = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$

$\quad \text{SSE} = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$

$\quad \text{SSR} = \text{SSTO} - \text{SSE} = \sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2$

coeffecient of determination $\quad R^2 = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 - \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$

$\quad R^2 = \frac{\text{SSTO} - \text{SSE}}{\text{SSTO}}$

$\quad R^2 = \frac{\text{SSR}}{\text{SSTO}}$

$\quad \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 - \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$

Functions

Quantile Function $Q(p)$ For a dataset consisting of n values that are ordered so that $x_1 \leq x_2 \leq \ldots \leq x_n$ and value p where $0 \leq p \leq 1$, let $i = \lfloor n \cdot p + 0.5 \rfloor$. Then the quantile function at p is:

$$Q(p) = \begin{cases} x_i & \lfloor n \cdot p + 0.5 \rfloor = n \cdot p + 0.5 \\ x_i + (n \cdot p - i + 0.5)(x_{i+1} - x_i) & \lfloor n \cdot p + 0.5 \rfloor \neq n \cdot p + 0.5 \end{cases}$$