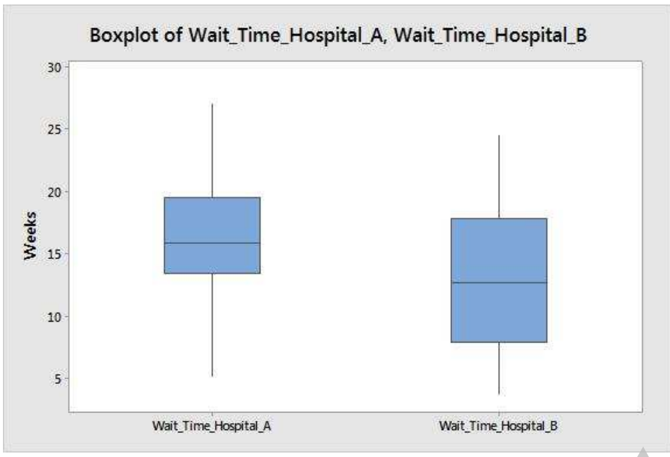
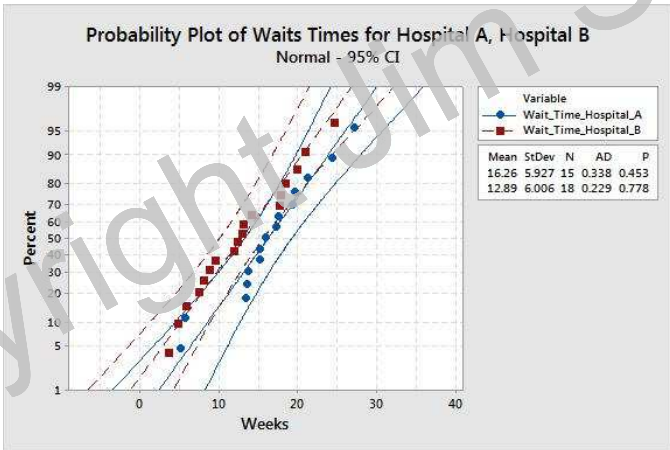
Sample Final Exam © J.Stallard 2017

1. The following data resulted from two random samples of patients waiting for knee replacement surgery. One sample was from a list of all patients who had knee surgery at Hospital A, the other was from a list of patients having knee surgery at Hospital B. These data are summarized below.

Variable	n	Mean	StDev	Q1	Median	Q3
Wait_Time_Hospital_A	15	16.26	5.93	13.50	15.91	19.56
Wait_Time_Hospital_B	18	12.89	6.01	7.93	12.69	17.92





Sum of the ranks of the wait times of the Hospital A data is 299. Sum of the ranks of the wait times of the Hospital B data is 262.

(a) How was this data collected? Identify the data collection method.

(b) A statistical test was applied to determine if these data came from populations that have the same variance. The value of the test statistic was found to be **0.176** with a P-value of 0.679. What is the name of this test? In addition, provide the probability statement which gives this P-value.

(c) Using your understanding of these data, does it appear that patients awaiting knee replacement surgery at Hospital A have to wait longer than patients who are to receive knee replacement at Hospital B? An appropriate statistical test was carried out producing a P-value of about 0.06. State the statistical hypotheses, the value of the test statistic, and the probability expression of the P-value for the appropriate statistical test.

2. The article "Factors Associated with Sexual Risk-Taking Behaviours Among Adolescents"[1] examined the relationship between gender and contraceptive use by sexually active teenagers. Each teenager in the study was classified according to gender and contraceptive use, the latter having three categories: rarely or never use, use sometimes, always use. The results are summarized in the table below.

	Contraceptive Use		
Gender	Rarely/Never	Sometimes/Most Times	Always
Female	210	190	400
Male	350	320	500

(a) Does the data indicate there is a relationship between gender and contraceptive use? Test using $\alpha = 0.05$.

(b) If there is no relationship between gender and contraceptive use, how many of the teenagers involved in the study would you expect to be female and rarely/never use contraceptives?

[1] *Journal of Marriage and Family*, 1994, pgs. 622 - 632.

3. In light of a transit strike, a large Canadian city is currently studying the feasibility of the private contracting of the city's transit service. Specifically, the city is interested in determining if 30% of the city's residents would support the contracting out of public transit.

If statistical sampling of 500 randomly selected residents of this city indicates that at least 30% of the city's residents would support the contracting out of public transit, the city will begin the process of accepting tenders from companies wishing to take over city transit services.

(a) State the null and alternative hypothesis of interest.

(b) If the probability of making a Type I error is regulated to be 5%, what is the maximum number of people in the sample of 500 that must support the privatization of public transit in order to reject the hypothesis in (a)?

(c) If only 25% of the city's residents would support the contracting out of public transit, what is the probability of concluding that at least 30% of the residents do support this?

(d) Suppose you were to test the statistical hypotheses

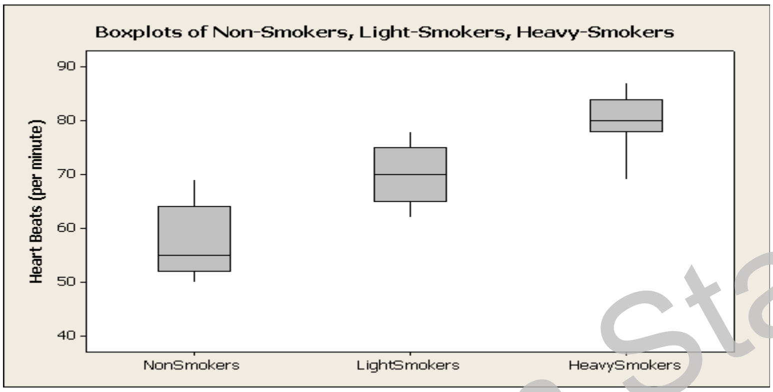
$$\mathrm{H}_0 : p = 0.30 \qquad\qquad \mathrm{H}_A : p = 0.25$$

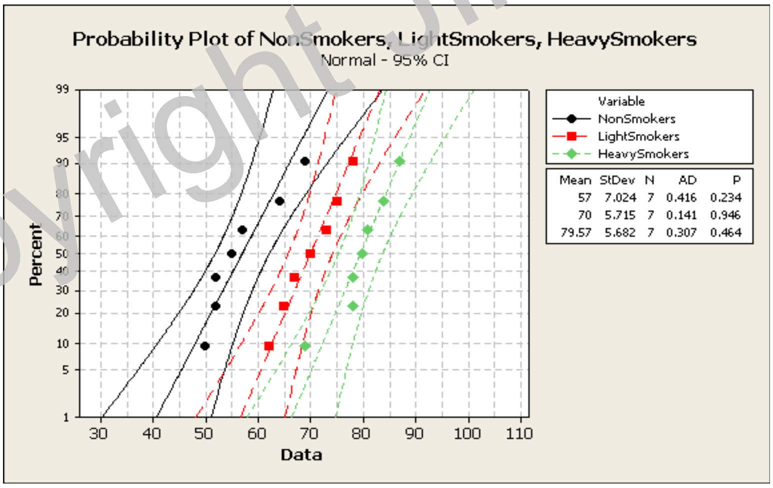
You wish to set $\alpha = 0.05$ and $\beta = 0.10$. How many city residents should you randomly select to test $\mathrm{H}_0 : p = 0.30$ against the alternative hypothesis $\mathrm{H}_A : p = 0.25$?

4. A researcher is to investigate the effect that smoking has on a person's heart rate while he/she is resting. She randomly selects seven people from a population of (1) Non-Smokers (2) Light-Smokers (few than 10 cigarettes/day) and (3) Heavy-Smokers (10 or more cigarettes/day) and records the heart-rate of each individual (or the number of times a person's heart-beats in a minute). The data and MINITAB output are provided below.

Non-Smokers	Light-Smokers	Heavy-Smokers
55	78	78
52	62	87
52	70	69
64	73	84
69	67	80
57	75	81
50	65	78

Boxplots each sample, as well as Normal Probability plots, are given.





A statistical model was applied to this data, producing the partial MINITAB output:

```
Source   DF       SS      MS       F       P
Factor        1796.9
Error
Total         2482.6
```

(a) A statistical test was conducted producing the given P-value of **0.0000096**. Testing at $\alpha = 0.05$, can you conclude from this data that, on average, the heart rate is the same for Non-Smokers, Light-Smokers, and Heavy-Smokers? Explain your answer.

(b) Refer to (a): Provide the hypotheses, value of test statistic, and P-value expression from the statistical test employed in (a).

(c) More MINITAB output is given below.

```
Level         N    Mean  StDev
NonSmokers    7  57.000  7.024
LightSmokers  7  70.000  5.715
HeavySmokers  7  79.571  5.682
```

Identify the difference(s), if any, with respect to the mean heart rate. Use $\alpha = 0.05$.

A portion of $q_{0.05}(k, n-k)$ table:

$n-k$:	$k=2$	$k=3$	$k=4$
16	3.00	3.65	4.05
17	2.98	3.63	4.02
18	2.97	3.61	4.00
19	2.96	3.59	3.98
20	2.95	3.58	3.96

(d) To check the assumption of equal variance, a certain statistical test was applied to the data producing a P-value of 0.831.

What can you conclude at $\alpha = 0.05$? **Indicate your answer in the box.**

(a) $\widetilde{\mu}_{Non} = \widetilde{\mu}_{Light} = \widetilde{\mu}_{Heavy}$
(b) $\widetilde{\mu}_{Non} \neq \widetilde{\mu}_{Light} \neq \widetilde{\mu}_{Heavy}$
(c) At least one of the $\widetilde{\mu}$'s is different
(d) $\sigma^2_{Non} = \sigma^2_{Light} = \sigma^2_{Heavy}$
(e) $\sigma^2_{Non} \neq \sigma^2_{Light} \neq \sigma^2_{Heavy}$
(f) At least one of the σ^2's is different
(g) $\mu_{Non} = \mu_{Light} = \mu_{Heavy}$
(h) $\mu_{Non} \neq \mu_{Light} \neq \mu_{Heavy}$
(i) At least one of the μ's is different

5. A specialist in traffic education wishes to test if three different methods of teaching defensive driving produce different results for men and women. Five people of each gender were randomly assigned to each of the three defensive driving classes: one was an eight-hour class, another involved two 4-hour classes, a third used two 2-hour classes. At the end of each program, a person was required to take a standardized test to examine their knowledge of defensive driving. The marks - out of 100 - are given, as well as relevant MINITAB output.

Gender	One 8-Hour session	Two 4-Hour sessions	Two 2-Hour sessions
Male	89	95	77
	96	87	78
	95	90	83
	90	91	78
	96	92	78
Female	88	87	80
	92	92	82
	98	91	79
	99	94	86
	91	93	88

(a) Create an interaction plot, with the 'session' acting as the x-axis and the gender acting as the y-axis.

(b) Consider the **partial** MINITAB output:

```
Two-way ANOVA: TestScore versus Gender, Session

Source       DF        SS       MS      F      P
Gender                20.83
Session              890.60
Interaction           24.07
Error                282.00
Total        29  1217.50

S = 3.428   R-Sq = 76.84%   R-Sq(adj) = 72.01%
```

Is there a statistically significant difference in the test scores between the three different methods of teaching defensive driving? The P-value of this test was found to be small enough to indicate the teaching method does effect how well a person did on the test. Sthat the statistical hypotheses that was tested and find the value of the test statistic used in this test.

6. A stock broker is interested in determining whether the yearly rate of return on Stock A, listed on the Toronto Stock Exchange (TSE), is linearly related to the yearly rate of return on the TSE Index. The following is a set of data showing historic yearly rates of return for Stock A and the TSE Index:

Year	Stock A	TSE Index
1	3.0%	5.0%
2	8.2%	13.5%
3	-6.0%	-12.5%
4	-9.5%	-20.2%
5	13.5%	17.5%
6	7.5%	14.5%

You are asked to model $RateOfReturn_A = \beta_0 + \beta_1 RateOfReturn_{TSE} + e_i$

$$\sum_{i=1}^{6} RateOfReturn_{TSE,i} = 17.80 \quad \sum_{i=1}^{6} RateOfReturn^2_{TSE,i} = 1288.0$$

Predictor	Coef	StDev	T	P
Constant	1.1308	0.6989		
TSE Inde	0.55703	0.04770		

S =   R-Sq =

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	383.27	383.27	136.36	0.000
Residual Error	4	11.24	2.81		
Total	5	394.51			

(a) State the simple linear regression equation based on the historical data. [2]

(b) Provide the value of the standard deviation of the regression.

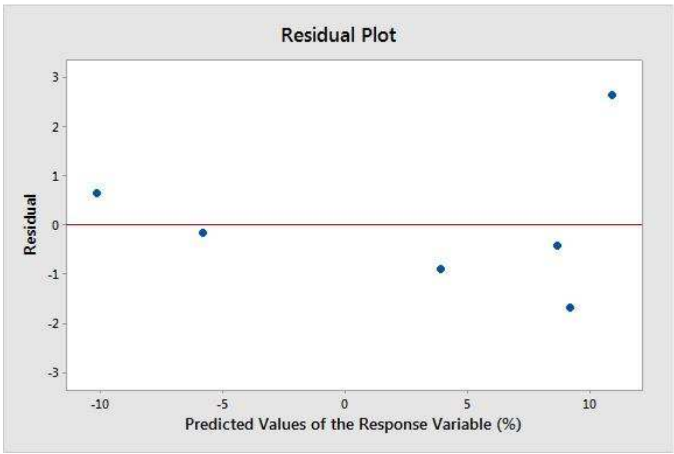
(c) Can the linear regression model in (a) be used to predict the yearly rate of return on Stock A? Test at $\alpha = 0.05$.

(d) Can one conclude that an increase in 1% in the TSE rate of return will translate into an average increase in the rate of return of Stock A by more than 0.40%? Test at $\alpha = 0.05$.

(e) What percentage of the variation in Stock A's annual rate of return is not explained by the model?

(f) An analyst estimates this year's TSE index to achieve a 5% rate of return. With 95% confidence, estimate this year's annual rate of return on Stock A. Interpret the meaning of this interval.

(g) Below is the residual plot. What condition does the residual plot 'visually check'? Does the condition appear to be satisfied? Address *both* questions, explain your answer in the area to the right of the residual plot. [2]



7. A manufacturer of warehouse-shelving units receives cement nails from six different suppliers. Random samples of the *same size* were drawn from each of the six suppliers and the number of defective cement nails from each supplier was observed. The number of defective nails is given below. (Note: n is does not need to be provided here)

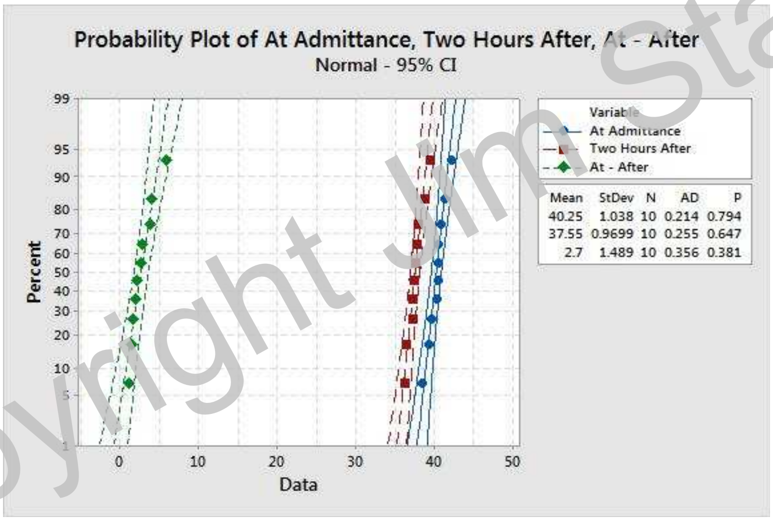
Supplier:	A	B	C	D	E	F
# of Defective Nails:	9	6	11	7	9	12

Do these data indicate that the proportion of defective nails is the same from each of the six suppliers? A statistical test produced a P-value of 0.72. State the statistical hypotheses, find the value of the test statistic, and give the probability expression for this P-value.

8. A drug designed to reduce fever is being tested for efficacy and side effects. Ten patients entering a hospital, all with high fever, are picked at random. Their body temperature (in degrees Centigrate) of each patient was measured at their admittance. Each was then given the drug, and their temperature recorded a few hours later. The data are given below.

Temperature At Admittance	Temperature a few hours After Admittance	Difference in Temperature
40.3	37.7	2.7
38.4	37.1	1.3
40.2	36.4	3.8
39.5	37.9	1.6
40.4	37.6	2.8
40.3	39.3	1.0
39.2	37.2	1.9
40.8	38.8	2.1
42.1	36.2	5.9
41.3	37.3	3.9

```
Variable          N    Mean   StDev
At Admittance    10  40.250   1.038
Two Hours After  10  37.550   0.970
At - After       10   2.700   1.489
```



A statistical test was carried out to see if this drug does lower temperature by 1 degree Celsius. This produced a P-value of 0.003. State the statistical hypotheses, find the value of the test statistic, and make a decision and conclusion.