Statistics for Final Year Engineering Students
SOLUTION TO: Final Class Test, Thursday May 26, 2011.
Lecturer: Dr. Yoni Nazarathy.

Duration: 40 Minutes.

Allowed: Non-Communicating calculators, double sided A4 reference sheet.

Not Allowed: Any other material, talking of any sort, looking side to side, passing items.
Students who violate this will be asked to leave immediately.

Write your answers CLEARLY in the answer boxes only. You may use spare paper supplied during the test – write your name on this sheet and hand it in along with the test.

A single sheet containing both a normal and an F distribution table is supplied.

The test is composed of 15 items. 7 points per item (Maximal grade: 105).

Good Luck.

1) An engineer is conducting an experiment to compare the durability of two tire types: Type X and type Y. 4 or 10 (two versions) tires of each type are put to a test and a unit of measure indicating the level of durability is determined for each of the tested tires. The engineer decides to conduct an ANOVA test and obtains a value of 5.42 for the F statistic. The engineer uses a confidence level of 95%.

Here is the critical value (different versions):

$$F_{0.95}^{(2-1=1, 4+4-2=6\ or\ 10+10-2=18)} = 5.99\ (4\ of\ each\ tire) or\ 4.41\ (10\ of\ each\ tire)$$

For each statement, indicate if true or false. And supply a **brief** explanation.

a) The engineer rejects H_0 and concludes that there is no difference between tire types.

True / False. Explanation: H_1 of ANOVA is that the means are NOT all the same. (Rejecting H_0 means choosing H_1).

b) The engineer rejects H_0 and concludes that the tire types have different durability levels with complete certainty.

True / False. Explanation: A result of an hypothesis test is never with certainty, there is always chance for error.

c) Knowing the value of the F statistic is not enough for deciding if to reject/not reject H_0 since the correlation coefficient between tires of type X and Y needs to also be computed.

True / False. Explanation: Correlation coefficient is not related to this ANOVA.

d) The engineer rejects H_0 and concludes the tire types differ in durability yet there is a 5% chance that the tire types do not have different durability levels.

True / False. In case of 4 tires False. In case of 10 tires True. Explanation: H_0 of ANOVA is that the means are the same, the type I error is 5%. Comparison to critical value in (a) yields the result.

e) The engineer does not reject H_0 .

True / False. In case of 4 tires True. In case of 10 tires False . Explanation:

2) Weights of items (measured in Kg) are denoted by $W_1, W_2, \dots, W_n$ and are assumed to be independent and follow the same continuous distribution with,

$$P(W_i > x) = \begin{cases} 2e^{-x} & a \le x \\ 1 & x < a \end{cases}.$$

a) Find the constant a.

$a = Log[2] = 0.69315$

Explanation:

For $a \le x$ $F(x)$ =1-$2e^{-x}$ and for $x < a$ $F(x)$ =1-$2e^{-x}$, solve for a: $F(a) = 0$.

Note: without assuming that the distribution is strictly continuous, there could also be a mass at a, yet in this subject we only discussed distributions that are either continuous or discrete.

b) Write (do not draw) the CDF of W_i, make sure you specify the value of the CDF for every x. (If you did not manage to solve the item above, leave your answer to this item and the next in terms of the constant a.)

$$F(x) = \begin{cases} 0 & x < Log[2] \\ 1 - 2e^{-x} & Log[2] \le x \end{cases}$$

c) Write (do not draw) the PDF of W_i.

$$f(x) = \begin{cases} 0 & x < Log[2] \\ 2e^{-x} & Log[2] \le x \end{cases}$$

d) An item is said to be "overweight" if it weighs more than 5 Kg. Assume $n = 100$. Write an expression (may involve a summation) for the probability of having less than 20 overweight items.

$$Probability\ of\ less\ than\ 20\ overweight = \sum_{k=0}^{19} \binom{100}{k} (2e^{-5})^k (1 - 2e^{-5})^{100-k}$$

Explanation: Chance of "success" is $p = 2e^{-5}$. Number overweight follows binomial distribution.

e) Let $\widetilde{W}_i$ denote the weight measured in **grams**. Write the PDF of $\widetilde{W}_i$ (if you did not answer (c), assume your result for (c) is $f(x)$ and try to write the result in terms of it).

$\tilde{f}(x) = \frac{1}{1000} 2e^{-x/1000}$

There was a lemma on this in the lecture notes (it is also easy to derive):

3) A shipping company is conducting a survey regarding container contents. 100 containers are picked at random and for each container the following quantities are recorded:
x_i- The weight of the container.
y_i- The value of the container contents.
I_i- Equals 1 if the container is privately owned and gets the value 0 if the container is owned by a company.

The following statistics are calculated: $\bar{x}$, $\bar{y}$ (sample means) S_x^2, S_y^2 (sample variances), $\hat{p}$ (sample proportion of privately owned containers) and $\sum_{i=1}^{100} x_i y_i$.

For each statement, indicate if true or false. And supply a **brief** explanation.

a) S_x^2 is an unbiased estimator of the variance of container weights.

True / False.
Explanation: The sample variance is unbiased.

b) The sample correlation coefficient of container weights and container values can be calculated based on the above statistics (without using any other information from the data).

True / False.
Explanation: By opening the square of the sample correlation coefficient, this can be achieved.

c) The sample correlation coefficient of container weights and container values is 0 because the sample is random and containers are assumed to be independent.

True / False.
Explanation ~~(no need for an exact formula if true):~~
Even though this is an independent sample each sampled container data gives two data bits (x and y) which may still be dependent.

d) A confidence interval for the proportion of privately owned containers with confidence level $1-\alpha$ can span a range (lowest value to upper value) of at most $\frac{z_{1-\frac{\alpha}{2}}}{10}$.

True / False.
Explanation:
We know that $x(1-x) \leq \frac{1}{4}$ for x in [0,1]. Plugging in the confidence interval formula yields the result.

e) $P(\hat{p}=1)$=$100p^{100}$, where p is the population proportion of privately owned containers.

True / False.
Explanation:
$P(\hat{p}=1)$= $100p(1-p)^{99}$