IE 121 – Spring 2002

Final exam

1. (20 points) In this problem, you need to select the correct answer.

   1. If the P-value for your test statistic is $P = .15$, then:

      (a) you would not reject H_0 for $\alpha = .05$.

      (b) you would reject H_0 for $\alpha = .05$.

      (c) you would reject H_0 for $\alpha = .10$.

      (d) your acceptance region has a lower limit of .15.

      (e) none of these

   2. Suppose you are going to test H_0: $m = m_0$, $H_1 : m \neq m_0$ using $\alpha = .05$. n and σ^2 are given. Find all values of $z = \frac{(\bar{x}-\mu)}{\sigma^2/\sqrt{n}}$ for which H_0 should be rejected.

      a) $z < -1.96$

      b) $z < -1.645$ or $z > 1.645$

      c) $-1.96 < z < 1.96$

      d) $z < -1.96$ or $z > 1.96$

      e) $-1.645 < z < 1.645$

   3. Which statement about the statistical process control is correct?

      a) A process is said to be operating in statistical control when chance causes are the only source of variation.

      b) A process that is operating in the presence of assignable cause is said to be in statistical control.

      c) The assignable causes include operator errors, defective raw materials and "background noise".

      d) All the points in a control chart should be in the control limits if there is no presence of assignable causes.

   4. Suppose we are interested in the mean weight of a certain type of a candy bar. We test 15 samples. The resulting sample average is 6.03 oz and the sample standard deviation is 1.47 oz. Then the 95% confidence interval for the mean weight is:

      a) $6.03 \pm 2.131 \cdot 0.38$

      b) $6.03 \pm 1.761 \cdot 0.38$

      c) $6.03 \pm 2.145 \cdot 0.38$

      d) $6.03 \pm 1.96 \cdot 0.38$

      e) None of these

5. The Central Limit Theorem is considered powerful in statistics because:

   a) It works for any sample size provided the population is normal.

   b) It works for any population distribution as long as the sample size is large enough.

   c) It works for any population distribution provided the population size is known.

   d) It works for any sample size provided the population distribution is known.

6. Which statement about the nonparametric statistics is incorrect?

   a) The Wilcoxon signed-rank test applies to the case of symmetric continuous distributions.

   b) A nonparametric procedure will be less efficient than the corresponding parametric procedure when the underlying population is normal.

   c) The distribution of the population will affect the probability of type II error of the sign test.

   d) The sign test is superior to the t-test if the population is approximately normal.

7. Suppose we did a study on how annual corn yield Y (bushel per acre) in Texas depends on the average annual rainfall R (inch) . This is the 10 years' data we collected. $Y=($ 101 102 105 115 117 120 123 123 128 133). $R=($28 28 29 32 33 33 34 35 36 37). We established the regression model, $\hat{Y}=12.2+3.22R$, to predict the future yield. Which of the following statement about the prediction is correct?

   a) If the average rainfall of next year is 30 in Texas, then the predicted average corn yield is 100.

   b) If the average rainfall of next year is 30 in PA, then the predicted average corn yield in PA is 108.8.

   c) If the average rainfall of next year is 300 in Texas, then the predicted average corn yield in Texas is 978.2.

   d) None of the above.

8. Which of the following cannot be done by a chi-square test?

   a) Test the hypothesis that the underlying distribution of a population is a uniform distribution.

   b) In a contingency table, test the hypothesis that the row-and-column methods of classification are independent.

   c) Test the hypothesis that the underlying distribution of a population is an exponential distribution.

   d) Wilcoxon Rank-Sum test.

9. Which of the following statements about the random error ϵ of the regression model $Y=\beta_0+\beta_1x_1+\beta_2x_2+\epsilon$ is correct?

   a) ϵ is a random variable with mean equal to β_0.

b) The variance of ϵ is smaller than variance of the response variable Y.

c) The distribution of ϵ is not necessarily symmetric.

d) When the value of the regressor x_1 or x_2 is larger, the variance of ϵ will be larger.

e) None of the above.

Answers: a-d-a-c-b-d-d-d-e

IE 121 – Spring 2002

Final exam

2. (10 points) Suppose the process is in control, and 3-sigma limits are in use on the $\bar{X}$ chart. The mean shifts by 1.5σ. The sample size is 4.

a) What is the probability that this shift will be detected in the first sample after the shift?

b) What would that probability be if 2-sigma limits were used?

Solution:

The corresponding material has not been covered this year.

IE 121 – Spring 2002

Final exam

3. (10 points) A quality control engineer wants to test whether the mean copper content in an alloy is equal to 10% at 0.05 significance level. A sample of size 10 reveals the following figures (in %):

9.99 10.06 10.05 10.07 10.05 9.97 10.08 10.10 9.96 10.06.

a) What would his conclusion be if he used the sign test?

b) What would it be if he used the Wilcoxon signed-rank test?

Solution:

a) $r^+ = 7$, $r^- = 3$, $r = \min\{7, 3\} = 3$

So $r > r^*_{0.05} = 1$, and therefore we conclude that we can't reject the hypothesis that the mean is equal to 10%.

b) $w^- = 6$, $w^+ = 49$ $w = \min\{6, 49\} = 6$

So $w < w^*_{0.05} = 8$, and we reject the above hypothesis.

IE 121 – Spring 2002

Final exam

4. (10 points) Suppose we are interested in fitting a simple linear regression model $Y = \beta_0 + \beta_1 x + \epsilon$ where β_0 is known. Assume that $E(\epsilon) = 0$ and $V(\epsilon) = \sigma^2$, as usual.

a) Find the least squares estimator of β_1.

b) What is the variance of the above estimator?

c) How does the variance from b) compare with the variance of $\hat{\beta}_1$ in case where both β_1 and β_0 are unknown?

Solution:

a) Sum of the squares of the deviations of the observations from the true regression line is

$$L = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

The least squares estimator of β_1 must satisfy

$$\frac{\partial L}{\partial \beta_1}\Big|_{\hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \hat{\beta}_1 x_i) x_i = 0,$$

or, simplifying,

$$\beta_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i.$$

Solving the above for $\hat{\beta}_1$, we obtain

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \beta_0 \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2}.$$

b) Using the fact the all Y_i's are mutually independent and that $V(Y_i) = \sigma^2$, we obtain

$$V(\hat{\beta}_1) = \frac{1}{(\sum_{i=1}^{n} x_i^2)^2} \sum_{i=1}^{n} x_i^2 \sigma^2 = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$$

c) If both β_1 and β_0 are unknown, the variance is

$$V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}},$$

which is always larger than what we have found in b)

IE 121 – Spring 2002

Final exam

5. (10 points) Let X be a random variable with the following probability distribution:

$$f(x) = \begin{cases} (\alpha+1)x^{\alpha}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the maximum likelihood estimator of α based on a random sample of size n.

Solution:

$$L(\alpha) = (\alpha+1)x_1^{\alpha} \dots (\alpha+1)x_n^{\alpha} = (\alpha+1)^n \prod_{i=1}^{n} x_i^{\alpha}$$

$$\ln L(\alpha) = n\ln(\alpha+1) + \alpha \sum_{i=1}^{n} \ln x_i$$

$$\frac{d \ln L(\alpha)}{d\alpha} = \frac{n}{\alpha+1} + \sum_{i=1}^{n} \ln x_i.$$

Equating the above to 0 and solving for α, we obtain

$$\alpha = -\frac{n}{\sum_{i=1}^{n} \ln x_i} - 1.$$

IE 121 – Spring 2002

Final exam

6. (10 points) A regression model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ has been fit using 20 observations. The calculated t-ratios $\hat{\beta}_i / se(\hat{\beta}_i)$, $i = 1, 2, 3$ are $t_0 = 1.45$ for β_1, $t_0 = 2.93$ for β_2 and $t_0 = 3.67$ for β_3.

a) Find P-values or the corresponding ranges for the t-statistics.

b) What conclusions can be made about the significance of the three regressors? Use $\alpha = 0.05$.

c) Find the corresponding 95% confidence intervals if the least squares estimates are: $\hat{\beta}_1 = 2.50$, $\hat{\beta}_2 = 3.40$ and $\hat{\beta}_3 = 4.20$.

Solution:

Multiple regression has not been covered this year.

IE 121 – Spring 2002

Final exam

7. (10 points) Suppose we want to test the hypothesis H_0: $\mu_1 = \mu_2$ versus its two-sided alternative. Both variances σ_1^2 and σ_2^2 are known. We can take a total of $n_1 + n_2 = N$ observations. How should they be allocated to the two populations if we want to maximize the power of the test $1 - \beta$ for any fixed difference $\mu_1 - \mu_2 \neq 0$?

Hint: the power of the test has a direct connection to the variance of the test statistic. So maximizing $1 - \beta$ is equivalent to...

Solution:

Maximizing the power of the test is equivalent to minimizing the variance of the test statistic – the difference of sample means. Thus we need to choose n_1 and n_2 so that the variance of the difference

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

is minimized which leads us to

$$n_1 = \frac{N\sigma_1}{\sigma_1 + \sigma_2}$$

and

$$n_2 = \frac{N\sigma_2}{\sigma_1 + \sigma_2}$$

IE 121 – Spring 2002

Final exam

8. (10 points) A sociologist claims that 15% of residents of a certain town are joggers. A poll of 300 residents was conducted and it was found that the P-value of the corresponding two-sided test is 0.12. What can we say about the number of joggers in the poll?

Solution:

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{x-45}{6.185}$$

On the other hand, we know that

$$2(1 - \Phi(|z_0|)) = 0.12,$$

and, therefore $|z_0| = 1.55$, which means that either $z_0 = 1.55$ or $z_0 = -1.55$.

In the first case, we have

$$\frac{x-45}{6.185} = 1.55,$$

and $x = 54.6 \approx 55$.

In the second case –

$$\frac{x-45}{6.185} = -1.55,$$

and $x = 35.4 \approx 35$.

Thus we conclude that, based on the information available, the number of joggers in the poll could be either 55 or 35.

IE 121 – Spring 2002

Final exam

9. (10 points) A linear regression model is used to relate a response y to 5 regressors on the basis of 25 sample points. The resulting coefficient of determination is $R^2 = 0.50$. What is your conclusion about the significance of regression based on the analysis of variance (f-test) with $\alpha = 0.05$?

Solution:

Multiple regression has not been covered.