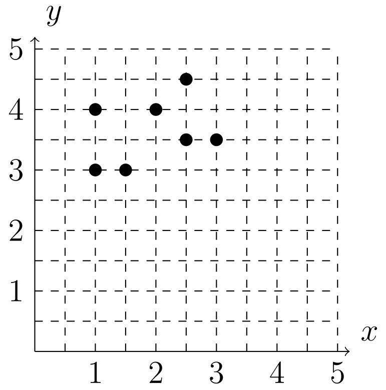
Name: ______________________________________

I.D. Number: ___________________

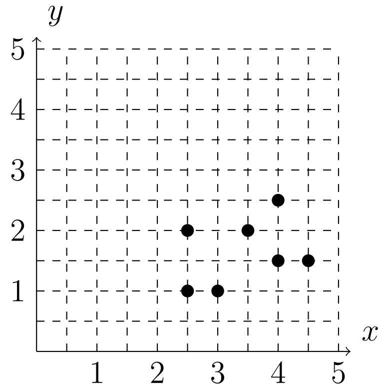
| | |
|---|---|
| 1 | /2 |
| 2 | /2 |
| 3 | /2 |
| 4 | /23 |
| 5 | /4 |
| 6 | /2 |

ECON 222
FINAL EXAM
MAY 21, 2013

1) Consider a bivariate data in which the relation between the quantities $x$ and $y$ can be described by: $y = \beta_0 + \beta_1 x + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$. The values of the parameters $\beta_0$ and $\beta_1$ are not know. We would like to find a confidence interval for $\beta_0$. For this purpose we take a sample of size 7. The scatter-plots of two possible samples are given below:



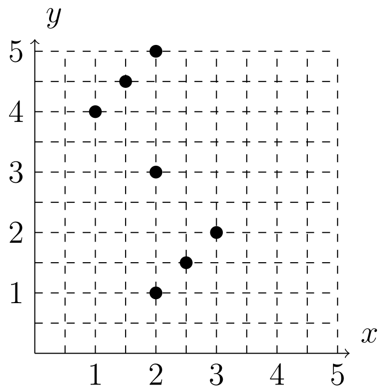

Scatter-plot of sample 1.



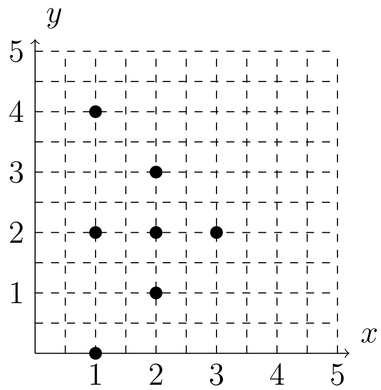

Scatter-plot of sample 1.

Would the widths of the confidence intervals (with same level of confidence) for $\beta_0$ obtained from these samples be different? If they would be different which sample (sample 1 or sample 2) would yield a narrower confidence interval (explain)?

2) Below you are given the scatter-plots of 2 different populations of $(x, y)$ pairs. Let $\rho_i$ be the population correlation coefficient of population $i$, $i = 1, 2$. What is the relation between the two population correlation coefficients (explain)?

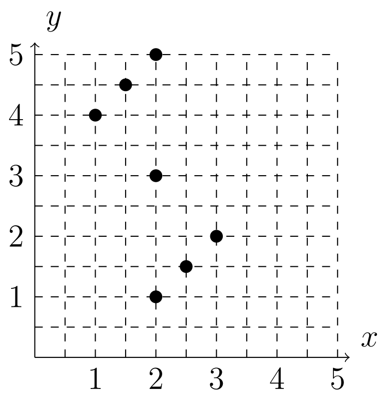


Scatter-plot of population 1.

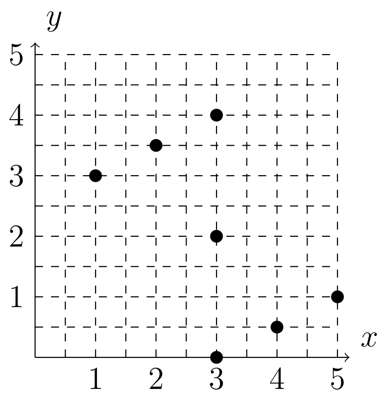


Scatter-plot of population 2.

3) Below you are given the scatter-plots of 2 different populations of $(x, y)$ pairs. Let $\rho_i$ be the population correlation coefficient of population $i$, $i = 1, 2$. What is the relation between the two population correlation coefficients (explain)?



Scatter-plot of population 1.



Scatter-plot of population 2.

4) We would like to test if there is a positive (linear) relationship between a persons income and the amount of life insurance held by that person. To conduct this test we take a random sample of size 10 for the population of all income earners and ask their annual income and life insurance they hold (measured in thousand USD). It is assumed that the relation between annual income, independent variable $x$, and life insurance they hold, dependent variable $y$, is given by: $y = \beta_0 + \beta_1 x + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$. The following is the part of the `EXCEL` output obtained from the `regression` tool of `EXCEL` :

| Regression Statistics | |
|---|---|
| $R$ Square | 0.5002 |
| Standard Error | 3.2397 |
| Observations | 10 |

| | Coefficients | Standard Error | $t$ Stat |
|---|---|---|---|
| Intercept | 10.6247 | 6.2255 | ____________ |
| $X$ Variable 1 | 0.1264 | 0.0447 | ____________ |

We are also given: $\bar{x} = 137.5$ and $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 5262.5$.

(a) As stated above, we would like test if there is a positive relation between annual income and amount of life insurance held:

(i) Write the hypotheses:

$H_0$ : ______________________________________________

$H_1$ : ______________________________________________

(ii) What is the test static that you would use to conduct this test:

(iii) What is the distribution of the test statistic:

(iv) What is the decision rule:

(v) What is the value of the test statistics:

(vi) What is your conclusion based on the sample:

(vii) What can you say about the corresponding $p$-value:

(b) Using this information find a 90% confidence interval for the slope coefficient ($\beta_1$)?

(c) Estimate the amount of life insurance held by an individual who earns 150 (thousand USD).

(d) Find a 90% confidence interval for the the amount of life insurance held by an individual who earns 150 (thousand USD).

(e) Find a 90% confidence interval for the the average amount of life insurance held by all individuals who earns 150 (thousand USD).

5) The relation between weakly income ($x$) and expenditure on food ($y$) is given by $y = \beta_0 + \beta_1 x + \epsilon$, where $\epsilon \sim N(0, 1)$ (assume that all the assumptions of simple linear regression holds.) We test $H_1 : \beta_1 < 1$ against $H_0 : \beta_1 \geq 1$ using the test statistics $(b_1 - 1)/\sigma_{b_1}$. The sample yields: $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 4$, $\bar{x} = 10$ and $\bar{y} = 2$. If $\beta_1 = 2$ what is the probability of a Type I error?

6) Answer one of the following questions:

(a) The correlation coefficient between $x$ and $y$ is equal to $-1$. You are also given that $\mu_x = 2$, $\mu_y = 5$, $\sigma_x = 6$, and $x = 4$ is paired with $y = 2$. What can you say about $\sigma_y$?

(b) Prove that, for any pair of random variables $X$ and $Y$,

$$\text{Var}[X + Y] = \text{Var}[X] + 2\text{Cov}(X, Y) + \text{Var}[Y] \ .$$

(You can think of the random variables $X$ and $Y$ as components of a bivariate data that have the same units.)

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

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

$$\hat{y}_i = b_1 x_i + b_0$$

$$s_\epsilon^2 = \frac{1}{n-2}\sum(y_i - \hat{y}_i)^2 = \frac{1}{n-2}\sum(y_i - (b_0 + b_1 x_i))^2 = \frac{\text{SSE}}{n-2}$$

$$\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2. \qquad \text{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2. \qquad \text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

$$y|_x \sim N(\beta_0 + \beta_1 x, \sigma^2).$$

$$\bar{y} \sim N\left(\beta_0 + \beta_1\bar{x}, \sigma^2/n\right)$$

$$b_0 \sim N\left(\beta_0, (\sigma^2 \sum x_i^2)/(n\sum(x_i - \bar{x})^2)\right)$$

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

$$b_0 + b_1 x \sim N\left(\beta_0 + \beta_1 x, \sigma^2\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)\right)$$

$$b_0 + b_1 x + \epsilon \sim N\left(\beta_0 + \beta_1 x, \sigma^2\left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)\right)$$

$$b_1 = r\sqrt{\frac{\sum(y_i - \bar{y})^2}{\sum(x_i - \bar{x})^2}} = r\frac{S_{yy}}{S_{xx}}$$

$$r^2 = \frac{SSR}{SST} = \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}.$$

$$F_{n_1,n_2,a} = \frac{1}{F_{n_2,n_1,1-a}}$$

Type I error: Rejecting $H_0$ when $H_0$ is true.

Type II error: Failing to reject $H_0$ when $H_0$ is false.