

Statistics XP 2015 - Quiz 3

NAME: _____

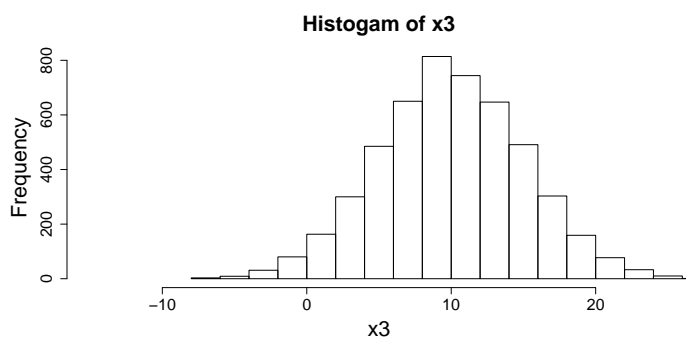
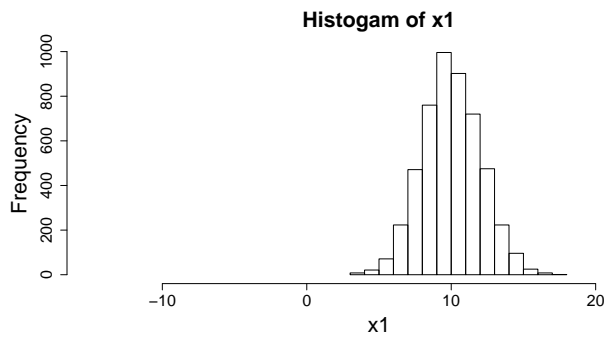
You have 20 minutes.

There are two questions, each part of each question is worth 2 points.

I pledge my honor that I have not violated the Honor Code during this examination.

SIGNATURE: _____

1 Question



Above are the histograms of x_1 , x_2 , and x_3 .

For each x the sample mean is either 0 or 10 and the sample standard deviation is either 2 or 5.

1.1

What is the sample mean of x_1 ?

1.2

What is the sample standard deviation of x_1 ?

1.3

What is the sample standard deviation of x_2 ?

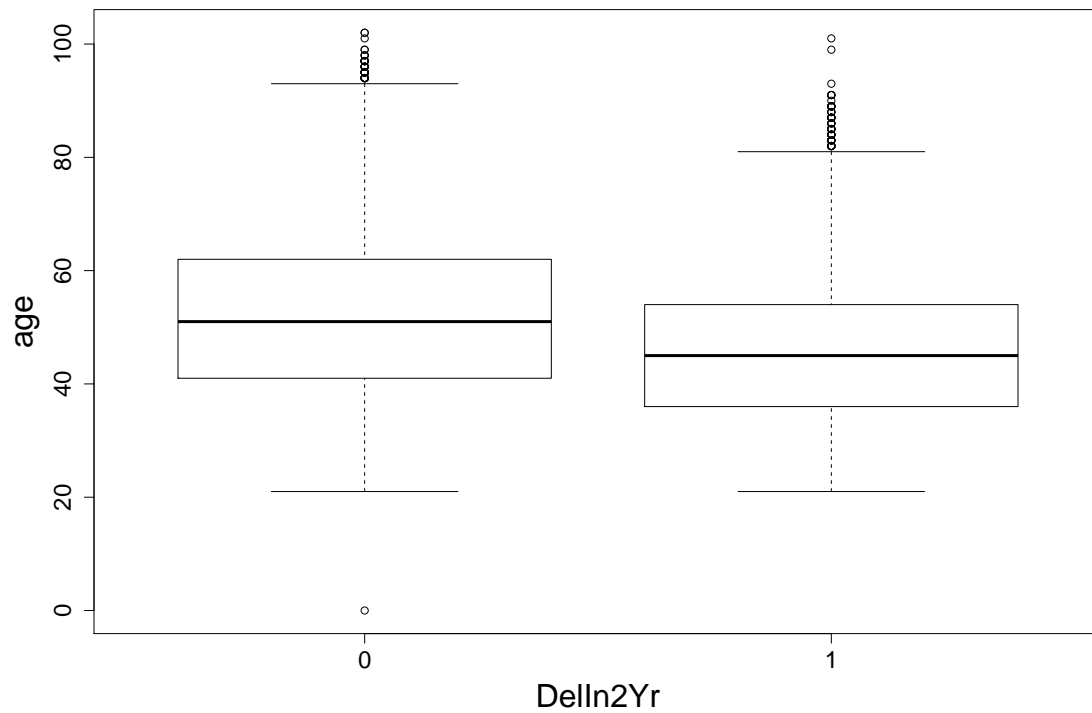
1.4

What is the sample variance of x_3 ?

1.5

Give an interval which should contain roughly 95% of the x_3 values.

2 Question



A company is trying to build a model to predict which customers will become delinquent in payments.

11,0521 observations were collected in which $y = \text{DelIn2Yr} = 1$ if an account is seriously delinquent in the next two years and 0 otherwise and $x = \text{age}$ is the age in years of the account holder.

Above are boxplots plotting $x = \text{age}$ versus $y = \text{DelIn2Yr}$.

2.1

What do the boxplots tell us about the relationship between $x = \text{age}$ versus $y = \text{DelIn2Yr}$?

```

Call:
glm(formula = DelIn2Yr ~ age, family = binomial, data = dd)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7599 -0.4122 -0.3485 -0.2942  2.9334

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.0946129  0.0437602  -25.01  <2e-16 ***
age          -0.0316261  0.0009107  -34.73  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 54266  on 110520  degrees of freedom
Residual deviance: 52982  on 110519  degrees of freedom
AIC: 52986

Number of Fisher Scoring iterations: 5

```

Above is the logistic regression output for the model

$$P(\text{DelIn2Yr} = 1 \mid \text{age}) = F(\beta_0 + \beta_1 \text{age})$$

where F is the logistic function $F(z) = \exp(z)/(1 + \exp(z))$.

2.2

Our estimate of the slope is $\hat{\beta}_1 = -0.0316261$.

What does the sign of $\hat{\beta}_1$ (the fact that it is negative) say about the relationship between age and delinquency?

2.3

Recall that $F(-3)$ is about .05.

For what age is the probability of delinquency .05?