Practice Exam # 1

STA 371G, Statistics and Modeling, Spring 2015

- Please answer all problems in the space provided on the exam.
- Read each question carefully and clearly present your answers.
- You must show all your work and give a complete explanation. No credit will be given for only the answer without an explanation/equation.
- The exam is closed-book. You are allowed one page of notes. You may use a calculator.

Z ~ N(0,1)

x	P(Z<x)	x	P(Z<x)	x	P(Z<x)	x	P(Z<x)
-3	0.0013	-1.5	0.0668	0	0.5	1.5	0.9332
-2.95	0.0016	-1.45	0.0735	0.05	0.5199	1.55	0.9394
-2.9	0.0019	-1.4	0.0808	0.1	0.5398	1.6	0.9452
-2.85	0.0022	-1.35	0.0885	0.15	0.5596	1.65	0.9505
-2.8	0.0026	-1.3	0.0968	0.2	0.5793	1.7	0.9554
-2.75	0.003	-1.25	0.1056	0.25	0.5987	1.75	0.9599
-2.7	0.0035	-1.2	0.1151	0.3	0.6179	1.8	0.9641
-2.65	0.004	-1.15	0.1251	0.35	0.6368	1.85	0.9678
-2.6	0.0047	-1.1	0.1357	0.4	0.6554	1.9	0.9713
-2.55	0.0054	-1.05	0.1469	0.45	0.6736	1.95	0.9744
-2.5	0.0062	-1	0.1587	0.5	0.6915	2	0.9772
-2.45	0.0071	-0.95	0.1711	0.55	0.7088	2.05	0.9798
-2.4	0.0082	-0.9	0.1841	0.6	0.7257	2.1	0.9821
-2.35	0.0094	-0.85	0.1977	0.65	0.7422	2.15	0.9842
-2.3	0.0107	-0.8	0.2119	0.7	0.758	2.2	0.9861
-2.25	0.0122	-0.75	0.2266	0.75	0.7734	2.25	0.9878
-2.2	0.0139	-0.7	0.242	0.8	0.7881	2.3	0.9893
-2.15	0.0158	-0.65	0.2578	0.85	0.8023	2.35	0.9906
-2.1	0.0179	-0.6	0.2743	0.9	0.8159	2.4	0.9918
-2.05	0.0202	-0.55	0.2912	0.95	0.8289	2.45	0.9929
-2	0.0228	-0.5	0.3085	1	0.8413	2.5	0.9938
-1.95	0.0256	-0.45	0.3264	1.05	0.8531	2.55	0.9946
-1.9	0.0287	-0.4	0.3446	1.1	0.8643	2.6	0.9953
-1.85	0.0322	-0.35	0.3632	1.15	0.8749	2.65	0.996
-1.8	0.0359	-0.3	0.3821	1.2	0.8849	2.7	0.9965
-1.75	0.0401	-0.25	0.4013	1.25	0.8944	2.75	0.997
-1.7	0.0446	-0.2	0.4207	1.3	0.9032	2.8	0.9974
-1.65	0.0495	-0.15	0.4404	1.35	0.9115	2.85	0.9978
-1.6	0.0548	-0.1	0.4602	1.4	0.9192	2.9	0.9981
-1.55	0.0606	-0.05	0.4801	1.45	0.9265	2.95	0.9984
-1.5	0.0668	0	0.5	1.5	0.9332	3	0.9987

Problem 1 (15 points)

A typical customer buys a random number (X) of polo shirts when he shops at a men's clothing store. The distribution of X is given by the following probability distribution: $P(X = 0) = 0.30$, $P(X = 1) = 0.30$, $P(X = 2) = 0.20$, $P(X = 3) = 0.10$, and $P(X = 4) = 0.10$.

(a) (3 points) Find the expected number of polo shirts bought by a customer when he visits this clothing store.

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i P(X = x_i) = 0 \times 0.30 + 1 \times 0.30 + 2 \times 0.20 + 3 \times 0.10 + 4 \times 0.10 = 1.40$$

(b) (3 points) Find the variance of X.

$$\begin{aligned}
\mathrm{Var}[X] &= \sum_{i=1}^{n} (x_i - \mathbb{E}[X])^2 P(X = x_i) \\
&= (0 - 1.40)^2 \times 0.30 + (1 - 1.40)^2 \times 0.30 + (2 - 1.40)^2 \times 0.20 \\
&\quad + (3 - 1.40)^2 \times 0.10 + (4 - 1.40)^2 \times 0.10 \\
&= 1.64
\end{aligned}$$

(c) (3 points) Assuming that each shirt costs \$50, let Y be the total amount of money (in dollars) spent by a customer when he visits this clothing store. Find the mean of Y.

Since $Y = 50X$, we have

$$\mathbb{E}[Y] = \mathbb{E}[50X] = 50\mathbb{E}[X] = 50 \times 1.40 = 70$$

(d) (3 points) Find the variance of Y.

$$\mathrm{Var}[Y] = \mathrm{Var}[50X] = 50^2\mathrm{Var}[X] = 2500 \times 1.64 = 4100$$

(e) (3 points) What's the probability that a customer's expenditure will be more than 1 standard deviation above the mean expenditure level?

The standard deviation of the random variable Y is $\sqrt{\mathrm{Var}[Y]} = 64.0$, the probability for Y to be larger than $70 + 64.0 = 134.0$ is

$$P(Y > 134.0) = P(X = 3) + P(X = 4) = 0.20$$

Problem 2 (15 points)

Suppose that 20% of the employees of a given corporation engage in physical exercise activities during the lunch hour. Moreover, assume that 60% of all employees are male, and 8% of all employees are males who engage in physical exercise activities during the lunch hour.

(a) (3 points) If we choose an employee at random from this corporation, what is the probability that this person is a *male* who does *not* engage in physical exercise activities during the lunch hour?

Let $M = 1$ if the person is male and $M = 0$ otherwise, and $A = 1$ if the person engages in physical exercise activities during the lunch hour and $A = 0$ otherwise. We have

$$P(A = 1) = 0.20, P(M = 1) = 0.60, P(A = 1, M = 1) = 0.08.$$

Thus

$$P(A = 0, M = 1) = P(M = 1) - p(A = 1, M = 1) = 0.60 - 0.08 = 52\%.$$

(b) (3 points) If we choose an employee at random from this corporation, what is the probability that this person is a *female* who engages in physical exercise activities during the lunch hour?

$$P(A = 1, M = 0) = P(A = 1) - P(A = 1, M = 1) = 0.20 - 0.08 = 12\%$$

(c) (3 points) If we choose an employee at random from this corporation, what is the probability that this person is a *female* who does *not* engage in physical exercise activities during the lunch hour?

$$P(A = 0, M = 0) = P(A = 0) - P(A = 0, M = 1) = (1 - 0.20) - 0.52 = 28\%$$

(d) (3 points) If we choose a *male* employee at random from this corporation, what is the probability that he engages in physical exercise activities during the lunch hour?

$$P(A = 1|M = 1) = \frac{P(A = 1, M = 1)}{P(M = 1)} = \frac{0.08}{0.6} = 13.3\%$$

(e) (3 points) If we choose at random a person who engages in physical exercise activities during the lunch hour, what is the probability that the person is a *female*?

$$P(M = 0|A = 1) = \frac{P(A = 1, M = 0)}{P(A = 1)} = \frac{0.12}{0.2} = 60\%$$

Problem 3 (25 points)

Amore Frozen Foods is reviewing the operation of their macaroni and cheese filling machine. Recall that the FDA approved weight control system for the company specifies that a 20 minute batch must be rejected if the average of 5 sample pies taken at the beginning of the batch run is less than 8.0 ounces.

Suppose the weight of the pies has a normal distribution with mean μ, where μ is the target fill rate, and a standard deviation of 0.22.

(a) (5 points) If Amore sets $\mu = 8.5$, what would be the 95% confidence interval of the weight of a randomly selected pie.

The 95% confidence interval would be

$$8.50 \pm 2 \times 0.22$$

(b) (5 points) What should the filling target be set to if Amore wants the probability to be 0.975 that a randomly selected pie will be no less than 8.0 ounces?

The weight of a pie is distributed as $X \sim \mathcal{N}(\mu, 0.22^2)$. Amore needs to set the target μ that makes

$$P(X \geq 8) = 0.975,$$

which means

$$P(X < 8) = 1 - 0.975 = 0.025.$$

Since

$$P(X < 8) = P(\frac{X - \mu}{0.22} < \frac{8 - \mu}{0.22}) = P(Z < \frac{8 - \mu}{0.22})$$

and $P(Z < -2) \approx 0.025$, we have $\frac{8-\mu}{0.22} = -2$. Thus the filling target should be set as

$$\mu = 8 + 2 \times 0.22 = 8.44.$$

(c) (5 points) What should the filling target be set to if Amore wants the probability to be 0.975 that a given batch will pass the FDA approved weight control system (i.e., a probability of 0.975 that the batch is accepted)?

For a sample of 5 pies, the sample mean $\overline{X}$ is distributed as

$$\overline{X} \sim \mathcal{N}\left(\mu, \left(\frac{0.22}{\sqrt{5}}\right)^2\right).$$

Amore needs to find a μ that makes

$$P(\overline{X} \geq 8) = 0.975,$$

which means

$$0.025 = P(\overline{X} < 8) = P\left(\frac{\overline{X} - \mu}{0.22/\sqrt{5}} < \frac{8 - \mu}{0.22/\sqrt{5}}\right) = P\left(Z < \frac{8 - \mu}{0.22/\sqrt{5}}\right).$$

Since $P(Z < -2) \approx 0.025$, we have $\frac{8-\mu}{0.22/\sqrt{5}} = -2$. Thus the filling target should be set as

$$\mu = 8 + 2 \times 0.22/\sqrt{5} = 8.20.$$

(d) (10 points) Suppose the company decides to set the target fill rate to $\mu = 8.20$. At this fill rate, the cost of ingredients per dozen pies is $\frac{8.20}{8.44} \times \$1.82 = \$1.77$. Also, the cost of packaging is \$0.62 per dozen pies and the cost of labor and overhead is \$0.56 per dozen pies. Therefore, the total cost is \$1.77 + \$0.62 + \$0.56 = \$2.95 per dozen pies.

A batch consists of 1,000 dozen pies (i.e. 12,000 pies). Suppose that if a batch is rejected then all 12,000 pies can be sold at the Thrift Store for \$3.60 per dozen pies. If a batch is not rejected then all 12,000 pies can be sold at wholesale for \$4.50 per dozen pies.

If the FDA approved weight control system is used, what is the expected profit per batch if the target fill rate is set to $\mu = 8.20$?

If a batch is rejected, the profit is

$$1000 \times (\$3.60 - \$2.95) = \$650.$$

If a batch is accepted, the profit is

$$1000 \times (\$4.50 - \$2.95) = \$1550.$$

Using the results in (c), we know that if we set $\mu = 8.20$, the probabilities for a given batch to be rejected and accepted are 0.025 and 0.975, respectively.

Thus the expected profit is

$$\mathbb{E}[\text{Profit}] = \$650 \times 0.025 + \$1550 \times 0.975 = \$1527.5$$

Problem 4: Match the Plots (10 points)

Below (Figure 1) are 4 different scatter plots of an outcome variable y versus predictor x followed by 4 four regression output summaries labeled A, B, C and D. Match the outputs with the plots. For example, if you believe Regression D matches Plot 1, you should draw a line to connect them.

Plot 1	B
Plot 2	C
Plot 3	A
Plot 4	D

Note: you can look at the R^2, intercept and slope to decide how the outputs and plots match each other.

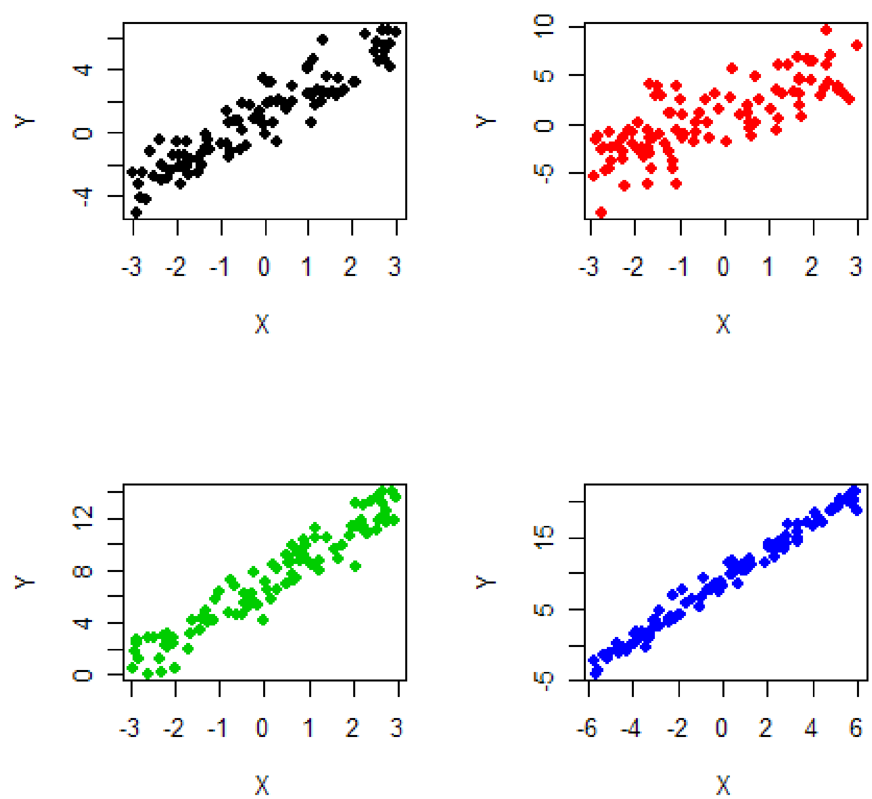

Figure 1: Scatter Plots

Regression A:

```
Coefficients:
                  Estimate   Std. Error
(Intercept)        7.03747    0.12302
(Slope)            2.18658    0.07801
---
Residual standard error: 1.226
R-Squared: 0.8891
```

Regression B:

```
Coefficients:
                  Estimate   Std. Error
(Intercept)        1.1491    0.1013
(Slope)            1.4896    0.0583

Residual standard error: 1.012
R-Squared: 0.8695
```

Regression C:

```
Coefficients:
                  Estimate   Std. Error
(Intercept)        1.2486    0.2053
(Slope)            1.5659    0.1119

Residual standard error: 2.052
R-Squared: 0.6666
```

Regression D:

```
Coefficients:
                  Estimate   Std. Error
(Intercept)        9.0225    0.0904
(Slope)            2.0718    0.0270
---
Residual standard error: 0.902
R-Squared: 0.9835
```

Problem 5 (20 points)

The table below shows the price (in thousands of dollars) for 6 houses (in thousands of square feet) recently sold in City Random. Let X denote Size and Y denote Price.

Sice (X)	0.80	0.90	1.00	1.10	1.40	1.40
Price (Y)	140	166	148	186	178	170

(a) (2 points) Calculate the sample means of X and Y.

$\bar{x} = 1.10$, $\bar{y} = 164.67$

(b) (2 points) Calculate the sample standard deviations of X and Y.

$s_x = 0.253$, $s_y = 17.603$

(c) (2 points) Calculate the sample covariance between X and Y.

$$\mathrm{Cov}(X,Y) = \frac{\sum_{i=1}^{6}(x_i - \bar{x})(y_i - \bar{y})}{6-1} = 2.88$$

(d) (2 points) Calculate the sample correlation between X and Y.

$$r_{xy} = \frac{\mathrm{Cov}(X,Y)}{s_x s_y} = \frac{2.88}{0.253 \times 17.603} = 0.647$$

(e) (2 points) Suppose we use simple linear regression to describe how the Price changes as a linear function of the Size. Calculate the least squares estimates of the intercept and slope.

$b_1 = r_{xy}\frac{s_y}{s_x} = 0.647 \times \frac{17.603}{0.253} = 45.0$, $\quad b_0 = \bar{y} - b_1\bar{x} = 115.2$

(f) (2 points) What are the units of the intercept and slope.

Intercept: 1000 dollars

Slope: 1000 dollars / 1000 square feet = dollar / square foot

(g) (4 points) Suppose we change the units of Size from 1000 square feet to 100 square meters (92.9 square meters= 1000 square feet), what would be the intercept and slope?

Intercept: 115.2 (in 1000 dollars)

Slope: 45 (in 1000 dollars / 1000 square foot) = 45 (in 1000 dollars / 92.9 square meters) = $45 \times \frac{100}{92.9}$ (in 1000 dollars / 100 square meters) = 48.4 (in 1000 dollars / 100 square meters)

(h) (2 points) Calculate the coefficient of determination R^2 and explain its meaning.

$R^2 = r_{xy}^2 = 0.42$ measures the proportion of variation in Y explained by X.

(i) (2 points) Based on this analysis, briefly describe your understanding of the relationship between Price and Size. The Price tends to increase as the Size increases. The Price increases by 45 thousand dollars as the Size increases by 1 thousand square feet.

Problem 6 (15 points)

Consider the regression model

$$\text{TI}_t = \beta_0 + \beta_1 \text{SP500}_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2),$$

where TI_t and SP500_t represent the returns in month t on Texas Instruments and the S&P 500, respectively. Suppose we know it is true that $\beta_0 = -0.0007$, $\beta_1 = 1.62$ and $\sigma = 0.091$.

(a) (5 points) Suppose the return on the S&P 500 is 3% in a given month, what's the expected return on TI in that month?

$$\mathbb{E}[\text{TI}_t \mid \text{SP500}_t = 3\%] = -0.0007 + 1.62 \times 3\% = 4.79\%$$

(b) (5 points) Suppose the return on the S&P 500 is -5% in a given month, what's the 95% prediction interval for the return on TI in that month?

$$\mathbb{E}[\text{TI}_t \mid \text{SP500}_t = -5\%] = -0.0007 - 1.62 \times 5\% = -8.17\%$$

95% Prediction Interval: $-0.0817 \pm 2 \times 0.091$

(c) (5 points) Suppose the return on the S&P 500 is 0% in a given month, what's the 95% prediction interval for the return on TI in that month?

$$\mathbb{E}[\text{TI}_t \mid \text{SP500}_t = -5\%] = -0.0007 + 1.62 \times 0\% = -0.07\%$$

95% Prediction Interval: $-0.0007 \pm 2 \times 0.091$