Statistics, Summer 2020 - Midterm

NAME: ____

Each part of each question is worth 2 points.
You have three hours to do the midterm.

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

SIGNATURE: ____

1 Question

The table below gives the joint distribution of X and Y.

		X = 0	X = 1
Y	0	.124	.056
	1	.216	.604

1.1

What is $P(X = 1, Y = 0)$?

1.2

What is $P(Y = 1)$?

1.3

What is $P(Y = 1 \,|\, X = 1)$?

1.4

Are X and Y independent ?

1.5

What is $E(X)$, the expected value of X?

1.6

What is $Var(X)$, the variance of X?

1.7

What is σ_X, the standard deviation of X?

1.8

Pick one:

(i) $Cov(X, Y) = 0$

(ii) $Cov(X, Y) > 0$

(iii) $Cov(X, Y) < 0$

2 Question

Suppose we model returns on Ford as $F \sim N(.12, 5.25)$ and returns on Tesla as $T \sim N(.14, 9.76)$ and believe $\text{Cov}(F, T) = 3.063$.

Let P represent the uncertain return on the porfolio which put 10% into Ford and 90% into Tesla.

$$P = .1\,F + .9T.$$

2.1

What is $E(P)$, the expected value of P ?

2.2

What is $Var(P)$, the variance of P ?

2.3

What is $sd(P)$, the standard deviation of P ?

2.4

Assuming P is normal, give an interval such that there is a 95% chance that P will end up in it?

2.5

Assuming P is normal, what is $P(P < 0)$?

3 Question

Suppose $R \sim N(3.5, 12.25)$ represents our beliefs about the return on an asset over the next period.

3.1

What is $E(R)$, the expected value of R ?

3.2

What is $Var(R)$, the variance of R ?

3.3

What is σ_R, the standard deviation of R ?

3.4

What is $P(-3.5 < R < 10.5)$?

3.5

What is $P(0 < R < 7)$?

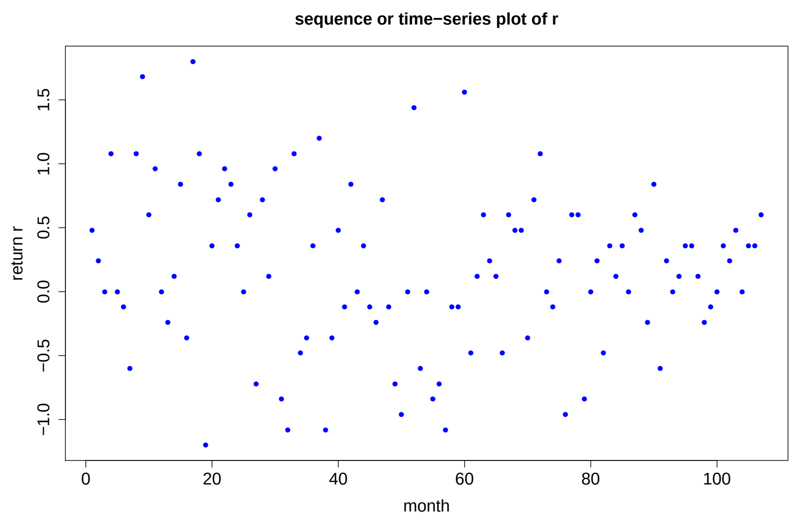
3.6

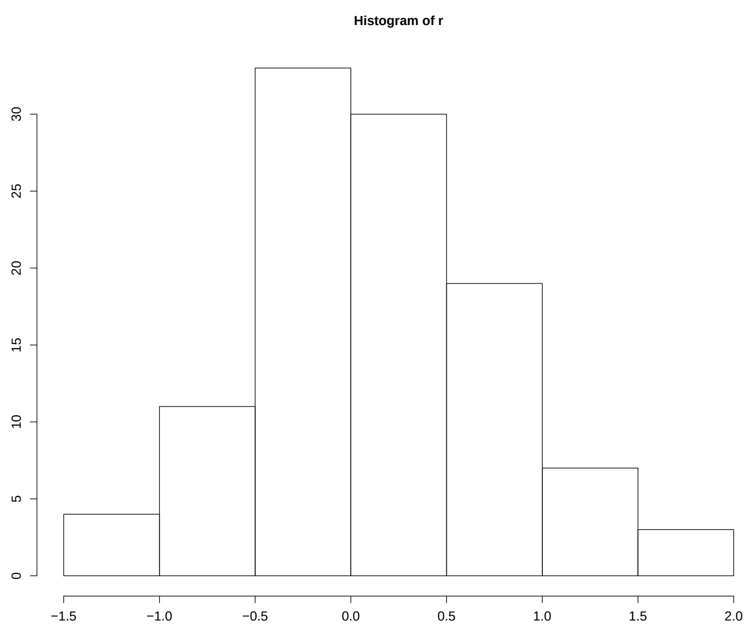
What is $P(0 < R)$?

4 Question

Let's look at the returns on a porfolio of assets from Denmark.
To make the numbers nicer, we multiply the monthly returns by 12.

```
> cd = read.csv("http://www.rob-mcculloch.org/data/conret.csv")
> r = cd$denmark*12
> mean(r)
[1] 0.1682243
> sd(r)
[1] 0.6289311
> length(r)
[1] 107
```





Above are the sample mean, sample standard deviation, and sample size.

Also, above, are the sequence plot (time-series plot) and the histogram.
To make the numbers nicer, we have multiplied the monthly returns by 12 to roughly annualize them.

So, the sample mean return is $\bar{r} = .168$, the sample standard deviation is $s_r = .629$ and the sample size is $n = 107$.

4.1

Based on the sequence plot and the histogram, do the returns "look" IID normal?

Discuss.

Assuming we are comfortable modeling the returns as IID normal, let's estimate the true mean μ, in our model

$$R_i \sim N(\mu, \sigma^2), \quad IID.$$

4.2

Using the sample mean as the estimate of the true mean, what is the associated standard error?

4.3

What is the 95% confidence interval for the true mean based on the 107 observations and the sample mean estimator?

4.4

As a practical matter, is the confidence interval big or small ?

4.5

As a practical matter, is there strong evidence that the true mean return is greater than 0?

4.6

What is the 68% confidence interval for the true mean based on the 107 observations and the sample mean estimator?

4.7

What is the 99% confidence interval for the true mean based on the 107 observations and the sample mean estimator?

4.8

Plugging in your estimates for μ and σ (sample mean and standard deviation), what is an interval such that there is a 95% chance the next return will be in it?

5 Question

In a recent poll, a random sample of 1,000 voters was taken from a *very* large population. Out of the 1,000 responders, 520 were in favor of proposition A and the rest were opposed.

Let p denote the true, population proportion of voters who are in favor.

5.1

What is $\hat{p}$, the usual estimate of p, based on our sample?

5.2

What is the standard error $\text{se}(\hat{p})$ associated with our estimate?

5.3

Give a 95% confidence interval for p.

5.4

Is it big or is it small?

5.5

Suppose you wanted a 95% confidence interval with a plus or minus of 1% ($\pm .01$).

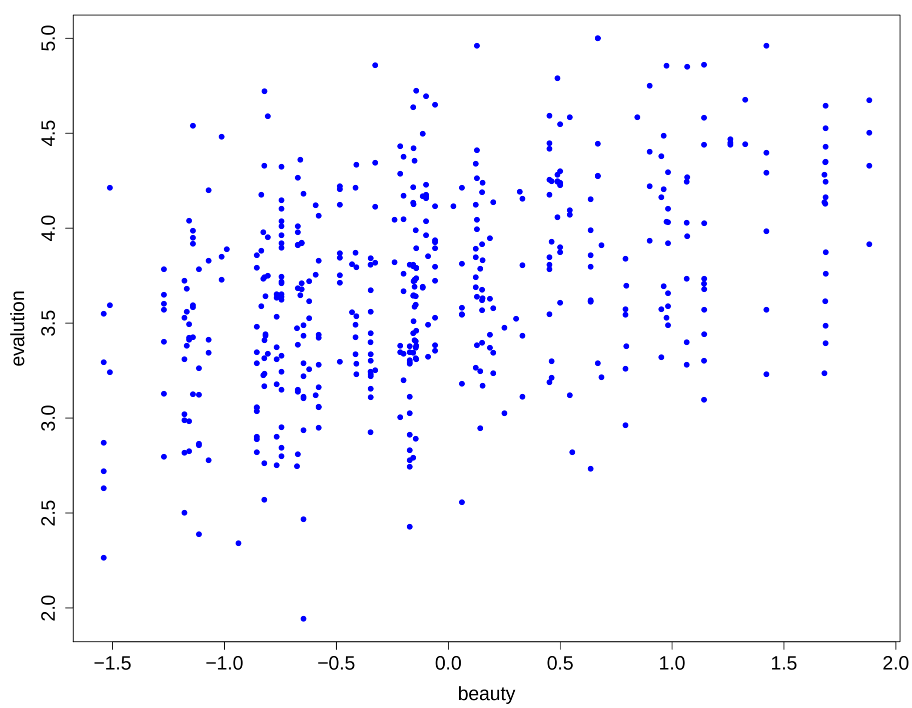
How big a sample size would you need?

6 Question

A researcher wanted to look into the relationship between the "beauty" of an instructor and their teaching ratings.

For each of $n = 463$ sections, the instructor's average evaluation was obtained as well as a measure of the "beauty" of the instructor.

Below is the scatter plot of x=beauty vs y=evaluation (average evaluation) for the 463 sections.
We are not sure how beauty was measured, but clearly it ranges from -1.5 to 2.0.



6.1

Based on the scatterplot, is the simple linear regression model a reasonable way to think about the relationship between x=beauty and y=evaluation?

Discuss.

Let's assume that we are comfortable with using the simple linear regression model

$$\text{evaluation} = \beta_0 + \beta_1 \,\text{beauty} + \epsilon, \quad \epsilon \sim N(0, \sigma^2).$$

Below is the regression output from the regression of y=evaluation on x=beauty.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.71340    ????    165.119   <2e-16 ***
beauty       0.27148    0.02837   9.569   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.4809 on 461 degrees of freedom
Multiple R-squared:  0.1657,Adjusted R-squared:  0.1639
F-statistic: 91.57 on 1 and 461 DF,  p-value: < 2.2e-16
```

6.2

Given beauty = 1.0, what is your prediction for evaluation?

6.3

Given beauty = 1.0, what is the plug-in prediction interval for evaluation?

6.4

What is the 95% confidence interval for the true slope β_1?

6.5

Test the null hypothesis that $\beta_1 = 0$ at level .05.

6.6

What is the standard error associated with the estimate of the true intercept β_0?

6.7

From the scatterplot, a large but plausible change in beauty would be from -1 to 1 resulting in a $\Delta x = 2$.

Would such a change result in a large change in the average rating?

6.8

Are x=beauty and y=evaluation related?

What do we learn from the regression analysis?

7 Question

Let's suppose the current (at the end of today) price of a stock is 100 (dollars).

Every day, the price either goes up by 2 or down by 1 (measured at the end of the trading day).
The probability it goes up is .6 (so the probability it goes down is .4).
Whether or not it goes up on a given day is independent of what happens on the other days.

That is, if Y_i is 1 if it goes up and 0 otherwise, then the Y_i are IID Bernoulli(.6).

Let P_i be the price at the end of day i, and today is day 0. So we know $P_0 = 100$.

7.1

What is $P(P_1 = 102)$.

That is, what is the probability that price at the end of day 1 (tomorrow) is 2 more than the price at the end of today (which is 100)?

7.2

What is $P(P_2 = 104)$.

7.3

What is $P(P_2 = 104 \mid P_1 = 102)$.

7.4

What is the distribution of P_1?
(List the possible values and their probabilites).

7.5

What is the distribution of P_2?
(List the possible values and their probabilites).

7.6

What is the distribution of P_3?
(List the possible values and their probabilites).

Let's suppose you own a call option on the stock with strike price $k = 101$ that expires at the end of the day three days from now.

This means that at the end of the day 3, you have the *option* of buying the stock at 101.

Let C be value of the option at the end of day 3.

So, if the price was 106, C would be 5 since you could buy it for 101 and then immediately sell for 106.

So, if the price was 97, C would be 0 since the option to buy at 101 is worthless.

$$C = \max(P_3 - k, 0)$$

where the max of two numbers is the biggest of the two.

7.7

At the end of day 3, what is the distribution of C?