Quiz 2, STATS 401 W18

In lab on 3/29 or 3/30

This document produces different random quizzes each time the source code generating it is run. The actual quiz will be a realization generated by this random process, or something similar.

Instructions. You have a time allowance of 50 minutes, though the quiz may take you much less time and you can leave lab once you are done. The quiz is closed book, and you are not allowed access to any notes. Any electronic devices in your possession must be turned off and remain in a bag on the floor.

Formulas

- The following formulas will be provided. To use these formulas properly, you need to make appropriate definitions of the necessary quantities.

(1) $\mathbf{b} = \left(\mathbb{X}^{\mathrm{T}}\mathbb{X}\right)^{-1}\mathbb{X}^{\mathrm{T}}\mathbf{y}$

(2) $\mathrm{Var}(\hat{\beta}) = \sigma^2\left(\mathbb{X}^{\mathrm{T}}\mathbb{X}\right)^{-1}$

(3) $\mathrm{Var}(\mathbb{A}\mathbf{Y}) = \mathbb{A}\mathrm{Var}(\mathbf{Y})\mathbb{A}^{\top}$

(4) $\mathrm{Var}(X) = \mathrm{E}\big[(X - \mathrm{E}[X])^2\big] = \mathrm{E}[X^2] - \big(\mathrm{E}[X]\big)^2$

(5) $\mathrm{Cov}(X, Y) = \mathrm{E}\big[\big(X - \mathrm{E}[X]\big)\big(Y - \mathrm{E}[Y]\big)\big] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$

(6) The binomial (n, p) distribution has mean np and variance $np(1-p)$.

From `?pnorm`:

```
pnorm(q, mean = 0, sd = 1)
qnorm(p, mean = 0, sd = 1)
q: vector of quantiles.
p: vector of probabilities.
```

Q1. Calculating means and variances, and making a normal approximation

Suppose that an analysis of a large dataset from another comparable university gave a coefficient of 0.03528 for the ACT variable when fitting a linear model using ACT score and high school rank. The admissions director is interested whether the difference could reasonably be chance variation due to having only a sample of 705 students, or whether the universities have differences beyond what can be explained by sample variation. Suppose that population value for this school is also 0.03528. Supposing the probability model in P1 and P2, and using a normal approximation, find an expression for the probability that the difference between the coefficient estimate for the data (0.03721) and the hypothetical true value (0.03528) is larger in magnitude than the observed value (0.03721-0.03528). Write your answer as a call to `pnorm()`. Your call to pnorm may involve specifying any necessary numerical calculations that you can't work out without access to a computer or calculator.

Let $X_1, X_2, \ldots, X_n$ be independent random variables each of which take the value 0 with probability 0.5, 1 with probability 0.25 and -1 with probability 0.25. Find the mean and variance of X_1. Use this to find the mean and variance of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Now suppose $n = 100$ and suppose that $\bar{X}$ is well approximated by a normal distribution. Find a number c such that $\mathrm{P}(-c < \bar{X} < c)$ is approximately 0.9. Write your answer as a call to `qnorm()`. Your call to qnorm may involve specifying any necessary numerical calculations that you can't work out without access to a computer or calculator.

Q2. Prediction using a linear model

Consider the following fitted linear model for ...

```
# lm1 <- lm()
# summary(lm1)$coef
```

We write the sample linear model as $\mathbf{y} = \mathbb{X}\mathbf{b} + \mathbf{e}$ where $\mathbb{X}$ is ... [I'M THINKING THERE WON'T BE TIME FOR WRITING OUT MANY MODELS. WE'LL HAVE THEM WRITE OUT THE PROBABILITY MODEL IN THE F-TEST QUESTION AND GIVE THEM THE MODEL WHEN NEEDED ELSEWHERE]

Suppose we're interested in predicting the value of a new observation at [explanatory variable values] together with a measure of our prediction uncertainty.

(a) Specify the values in a row matrix $\mathbf{x}^*$ so that $\mathbf{y}^* = \mathbf{x}^*\mathbf{b}$ gives a least squares prediction of the new observation.

(b) Give a matrix

Q3. Comparing means using a linear model

This question will be based on HW7. It will involve either confidence interval construction or making a hypothesis testing when comparing means of two samples in the context of a linear model.

[A confidence interval question]

[A hypothesis testing question]

Q4. Making and interpreting an F test

The following model was fitted to a dataset ... Write out the null and alternative hypothesis, the test statistic, the p-value and your conclusion.
