1 Writing Hypotheses

1. The General Social Survey collects data on demographics, education, and work, among many other characteristics of US residents. A data anlyst examines the data from 2010 and finds those with a college degree worked 41.8 hours on average, those without a college degree worked 39.4 hours. Write hypotheses that test whether there is any significant difference between average hours of work of those who have a college degree and those who do not.

2. According to a New York Times article sixty-percent of pregnancy related deaths can be prevented. Based on the data provided by Center for Disease Control on pregnancy related deaths in the US, the writer reports that black women were more likely than white women to suffer a pregnancy-related death.Write hypotheses that can test this claim.

3. A statistics professor thinks that talking to students about self-confidence will help them improve their self confidence. She gives students a survey about self-confidence and then uses the scores to calculate a self-confidence score of each student. Then she talks to students about self-confidence and then gives them the survey again. Write hypotheses that tests the statistics professors question.

4. A data scientist is A/B testing. In one version of a web page (version A) users spend on average 3.18 minutes on the web page. In the other version of the web page (version B) they spend on average 2.97 minutes. Write hypotheses to test whether version A leads to users spending more time on the web page than version B.

5. The following is an excerpt from a research study

   Evaluation of children with Autism Spectrum Disorders (ASD) is crucial to clinical diagnosis and educational intervention. The traditional evaluation methods based on questionnaires and scales rely on the experience and expertise of the evaluator, are time-consuming and clinically demanding. Computer games can provide an objective, motiving and safe way for evaluating and reflecting children's development. Therefore, the study aimed to investigate a technology-based method using computer games to evaluate children with ASD. The performance of 40 children with ASD and 51 aged-matched typically developing (TD) children was compared. We found: 1) The completion ratio for children with ASD was lower than TD children for the tasks in most of the games.

Write out hypotheses that would test this claim for a single game.

Under certain conditions:

$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

$(\bar{x}_1 - \bar{x}_2) \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

$\hat{p} \sim N(p, \frac{p(1-p)}{n})$

$(\hat{p}_1 - \hat{p}_2) \sim N((p_1 - p_2), \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2})$

2 Hypothesis Testing

At the beginning of the quarter I had asked you how many miles you are away from home. The following is based on a sample of the data.

```
## # A tibble: 1 x 3
##   mean_miles sd_miles     n
##        <dbl>    <dbl> <int>
## 1      1157.    2311.    40
```

Set and test hypotheses whether the average distance to home is more than 1000 miles for everyone in our class.