STAT 512 su 2021 hw 7

1. In a study of offspring sex-ratios in the Swedish-born population in [3], it is reported that among the 2,059,372 first-born children during the time period from 1932 until 2013, 1,058,701 were males, and among the 1,699,793 second-born children during the same time period, 874,369 were males.
   (a) Build a 95% confidence interval for the difference $p_1 - p_2$, where p_1 is the proportion of first-born children who are males and p_2 is the proportion of second-born children who are males.
   (b) Report the margin of error for the confidence interval.
   (c) Do you believe there is a difference between the proportion of males among first-born children and among second-born children? Base your answer on the data.

2. In [2], continuing with the Swedish theme, a data set was analyzed in which the number of traffic accidents in a day on Swedish roads was measured when a speed limit was and was not enforced. The measurements were taken on several days during the years 1961 and 1962. Read in the data by installing the R package `MASS` and executing the following commands:

```
library(MASS)
X <- Traffic$y[Traffic$limit=="no"]
Y <- Traffic$y[Traffic$limit=="yes"]
```

   (a) Now construct a 95% confidence interval for the difference in the mean number of traffic accidents under enforcement and non-enforcement of a speed limit (non-enforcement minus enforcement). Assume non-Normality of the population distributions.
   (b) Give an interpretation of your confidence interval. What is your recommendation about speed limits?

3. The number of insects on the leaves of some plants were counted after the application of different pesticides. The data for pesticides "A" and "B", taken from [1], can be read into R with:

```
data("InsectSprays")
X <- InsectSprays$count[InsectSprays$spray=="A"]
Y <- InsectSprays$count[InsectSprays$spray=="B"]
```

   (a) We know that these random samples were *not* drawn from Normal distributions. Explain why we know this without having to do any analysis.
   (b) Treating the random samples as though they *were* drawn from Normal populations, give a 95% confidence interval for the ratio σ_2^2/σ_1^2, where σ_2^2 is the variance for pesticide "B" and σ_1^2 is the variance for pesticide "A".
   (c) Comment on whether there is evidence to conclude that the two variances are unequal.
   (d) Construct a 95% confidence interval for the difference $\mu_1 - \mu_2$, where μ_1 and μ_2 are the mean numbers of insects after the application of pesticides "A" and "B", respectively.
   (e) Give an interpretation of the confidence interval.

4. Let $X_1, \dots, X_{n_1} \overset{\text{iid}}{\sim} \text{Normal}(\mu_1, \sigma^2)$ and $Y_1, \dots, Y_{n_2} \overset{\text{iid}}{\sim} \text{Normal}(\mu_2, \sigma^2)$ be independent random samples (note that the population variances are both equal to σ^2) with sample variances S_1^2 and S_2^2, respectively.

(a) Show that

$$S^2_{\text{pooled}} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

is an unbiased estimator of σ^2.

(b) Use the fact that

$$\frac{(n_1 + n_2 - 2)S^2_{\text{pooled}}}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$$

to derive the upper and lower bounds of a $(1 - \alpha)100\%$ confidence interval for σ^2.

5. Consider estimating a population proportion p.
   (a) What is the most conservative sample size (erring on the large size) required in order to build a 95% confidence interval for a population proportion p with a margin of error of at most 2%?
   (b) What about with a margin of error of at most 1%?
   (c) What about with a margin of error of at most 0.5%?
   (d) If you quadruple the sample size, what happens to the width of the confidence interval?

6. Researchers are interested in comparing the means μ_1 and μ_2 of two populations. A pilot study has suggested that the standard deviation σ_1 of the first population is three times larger than the standard deviation σ_2 of the second population.
   (a) The researchers have the resources to sample a total of $1{,}000$ observations from the two populations. Find the number of observations n_1 which should be drawn from population 1 and the number of observations n_2 which should be drawn from population 2 such that the width of a confidence interval for $\mu_1 - \mu_2$ is minimized.
   (b) Suppose that the pilot study suggested $\sigma_1 \approx 1$. Recommend sample sizes n_1 and n_2 under which the margin of error of a 98% confidence interval for $\mu_1 - \mu_2$ will be less than 0.10.

Optional (do not turn in) problems for additional study from Wackerly, Mendenhall, Scheaffer, 7th Ed.:

- 8.61, 8.62, 8.64
- 8.70, 8.71, 8.74, 8.76
- 8.82, 8.83

References

[1] Geoffrey Beall. The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. *Biometrika*, 32(3/4):243–262, 1942.

[2] Åke Svensson. On a goodness-of-fit test for multiplicative poisson models. *The Annals of Statistics*, pages 697–704, 1981.

[3] Brendan P Zietsch, Hasse Walum, Paul Lichtenstein, Karin JH Verweij, and Ralf Kuja-Halkola. No genetic contribution to variation in human offspring sex ratio: a total population study of 4.7 million births. *Proceedings of the Royal Society B*, 287(1921):20192849, 2020.