Formula sheet for final exam

$$\bar{x} = \frac{\sum x_i}{n}, \quad s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}, \quad IQR = Q_3 - Q_1, \qquad z = \frac{x_i - \bar{x}}{s} \text{ or } z = \frac{x_i - \mu}{\sigma}$$

For a finite probability distribution

$$\mu = \sum kP(X = k), \quad \sigma^2 = \sum (k - \mu)^2 P(X = k).$$

A special case is the binomial with parameters n and p. For this we have

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \mu = np, \quad \sigma = \sqrt{np(1-p)}.$$

The central limit theorem states that if $x_1, x_2, \ldots, x_n$ is a random sample from a population with mean μ and standard deviation σ, then the sample mean is approximately normal with mean μ and standard deviation $\sigma/\sqrt{n}$.

Under assumptions, $(1-\alpha)100\%$ CIs for the population proportion p based on $\widehat{p}$ and the population mean μ based on $\bar{x}$ are respectively

$$\widehat{p} \pm z^* \sqrt{\widehat{p}(1-\widehat{p})/n} \quad \text{and} \quad \bar{x} \pm t^* s/\sqrt{n},$$

where t^* and z^* and related to α by

$$P(-z^* \leq Z \leq z^*) = 1 - \alpha, \quad P(-t^* \leq T_{n-1} \leq t^*) = 1 - \alpha.$$

Test statistics The following test statistics may prove useful. Ones labeled Z have a normal sampling distribution under the proper assumptions, ones labeled T have a t-distribution. The degrees of freedom appear beside it.

$$Z = \frac{\widehat{p} - p}{\sqrt{p(1-p)/n}} \qquad Z = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})}\sqrt{1/n_1 + 1/n_2}}$$

$$T = \frac{\text{observed} - \text{expected}}{\mathsf{SE}} \qquad T = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad n-1 \text{ d.f.}$$

$$T = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}}, \quad n_1 + n_2 - 2 \text{ d.f.} \qquad T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \quad \text{smaller of } n_1 - 1, n_2 - 1$$

(Use the pooled standard deviation when you assume $\sigma_1 = \sigma_2$, otherwise, use the other value to find SE.)

Regression The Pearson correlation and regression coefficients are given by

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}, \qquad \hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

This test statistic is used for testing the β's:

$$T = \frac{\hat{\beta}_i - \beta_i}{\mathsf{SE}} \quad n - 2 \text{ d.f..}$$

Finding p-values You will be asked to find a p-value for many of these questions. You have a copy of the normal table and for the tail of the t distribution attached to the exam As well you may use the computer (hint: `pnorm()` or `pt()`). If you are using a table, you may not be able to give the exact p-value. In this case, use the table to give an upper and lower value for the p-value. For instance, something such as $0.10 > p > 0.05$.