Statistics 431: Statistical Inference
Facts and Formulas for the Final Exam

Probability foundations

The normal distribution and its samples

- The probability density function of a $N(\mu, \sigma^2)$ rv is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) .$$

- The population mean is μ; the population SD is σ.
- For a sample $X_1, \ldots, X_n$ of size n from a normal population,
  - The sample estimate of μ is the sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

  - The sample estimate of σ^2 is the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\bar{X}\right)^2\right] .$$

- The distribution of the normal sample mean: $\bar{X} \sim N(\mu, \sigma^2/n)$. The SD of $\bar{X}$, $\sigma/\sqrt{n}$, is also called the standard error (SE) of $\bar{X}$. We estimate it as $S/\sqrt{n}$.
- $\bar{X}$ and S^2 are independent as rvs. The distribution of the sample variance: $(n-1)S^2 \sim \chi^2_{n-1}$.
- The sample histogram and the normal quantile plot are two graphical tools to judge whether a sample comes from an approximately normal population. Prefer the quantile plot.

The binomial distribution and its samples

- Let X count the number of "successes" in n independent Bernoulli trials, each with probability p of "success." Then X has the binomial distribution, $X \sim \text{Bin}(n, p)$. The probability mass function of a binomial rv is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} ,\ k = 0, \ldots, n .$$

- The population mean is np; the population SD is $\sqrt{np(1-p)}$.
- For a binomial rv X based on a Bernoulli sample $Z_1, \ldots, Z_n$ from a population, the estimate of p is $\hat{p} = X/n$.
- The SD of $\hat{p}$ is $\sqrt{p(1-p)/n}$; it is also called the SE of $\hat{p}$. We estimate it as $\sqrt{\hat{p}(1-\hat{p})/n}$.
- For large n, the distribution of $\hat{p}$ is approximately $N(p, p(1-p)/n)$.

Chapter 7: Confidence intervals

One sample mean

- A $100\gamma\% = 100(1-\alpha)\%$ confidence interval (CI) for an unknown population mean μ has the general form

  $$\bar{X} \pm C^* \cdot \frac{\sigma^*}{\sqrt{n}} = \left(\bar{X} - C^* \cdot \frac{\sigma^*}{\sqrt{n}},\ \bar{X} + C^* \cdot \frac{\sigma^*}{\sqrt{n}}\right) ,$$

  where C^* is an appropriate upper quantile and σ^* is an appropriate population SD or estimate thereof. The meaning of the confidence statement is that $P(\mu \in \text{Interval}) = \gamma$, at least approximately. The important situations are:

  - σ known and either population normal or n large: $\sigma^* = \sigma$ and $C^* = z_{\alpha/2}$, the $(\alpha/2)$ upper quantile of the standard normal.
  - σ unknown and n large: $\sigma^* = S$ and $C^* = z_{\alpha/2}$.
  - σ unknown and population normal: $\sigma^* = S$ and $C^* = t_{\alpha/2;n-1}$, the $(\alpha/2)$ upper quantile of the t distribution with $n-1$ degrees of freedom (df).

- The sample size needed to get an interval of width w is (approximately)

  $$n(w) = \left(2z_{\alpha/2}\frac{\sigma}{w}\right)^2 ,$$

  rounded up to the nearest integer. When σ is unknown, use an estimate from previous experience or from the corresponding value of S in a pilot experiment.

One population proportion

- When n is large (say, $n\hat{p}(1-\hat{p}) > 20$) and $\hat{p}$ is not too near 0 or 1, you can use the classical large-sample CI formula,

  $$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} .$$

- Otherwise, use Wilson's formula

$$\frac{\hat{p} + z_{\alpha/2}^2/2n}{1 + z_{\alpha/2}^2/n} \pm z_{\alpha/2} \cdot \frac{\sqrt{\hat{p}(1-\hat{p})/n + z_{\alpha/2}^2/4n^2}}{1 + z_{\alpha/2}^2/n} .$$

- When in doubt, use Wilson's formula – it is almost equivalent to the large-sample formula on large samples, but is better on small samples.

- The sample size needed to get an interval of width w is

$$n(w) = \left(2z_{\alpha/2} \frac{\sqrt{\tilde{p}(1-\tilde{p})}}{w} \right)^2 ,$$

rounded up to the nearest integer. Here $\tilde{p}$ is a prior guess for p. A worst-case (possibly too large) n is obtained by letting $\tilde{p} = 1/2$.

One-sided confidence confidence interval (aka confidence bound)

- A $100(1-\alpha)\%$ upper confidence bound results from replacing $\pm$ in the above by $+$, and $z_{\alpha/2}$ by z_α (or $t_{\alpha/2;n-1}$ by $t_{\alpha;n-1}$). The other end of the interval is $-\infty$.

- A $100(1-\alpha)\%$ lower confidence bound results from replacing $\pm$ in the above by $-$, and $z_{\alpha/2}$ by z_α (or $t_{\alpha/2;n-1}$ by $t_{\alpha;n-1}$). The other end of the interval is $+\infty$.

Prediction intervals

- Let Y be a single future observation from a normal population distribution. A $100(1-\alpha)\%$ prediction interval for Y takes the form $\bar{X} \pm C^* \cdot \sigma^* \sqrt{1+1/n}$, with C^* and σ^* as above. The interval has the property that $P(Y \in \text{Interval}) \approx 1-\alpha$.

Chapter 8: Hypothesis tests

General theory

- Tests involve a null hypothesis H_0 and an alternative hypothesis H_A. A typical case tests $H_0 : \mu = \mu_0$ vs $H_A : \mu \neq \mu_0$ for an unknown mean parameter μ. This is a two-sided test. A one-sided test looks like $H_0 : \mu \leq \mu_0$ vs $H_A : \mu > \mu_0$. To conduct the test, we need a test statistic T and a rejection region like $|T| > c$ or $T > c$. Here c is the critical value.

- Suppose we are testing $H_0 : \mu \leq \mu_0$ vs $H_A : \mu > \mu_0$. Then

$$P_{\mu \in H_0}(\text{Reject } H_0) = P(\text{Type I error}) \text{ for this } \mu ,$$

and

$$P_{\mu \in H_A}(\text{Do not reject } H_0) = P(\text{Type II error}) = \beta \text{ for this } \mu \; .$$

The significance level α is the probability of a Type I error at the boundary value μ_0. The power of the test is $P_{\mu \in H_A}(\text{Reject } H_0)) = 1 - \beta$.

- If we observe the test statistic value $T = t$, the p-value is the smallest α at which we can reject H_0 using t. If you know the p-value of a test, you know the outcome for *every* level α:

$$p\text{-value} < \alpha \;\Rightarrow\; \text{reject} \; ; \;\; p\text{-value} \geq \alpha \;\Rightarrow\; \text{do not reject} \; .$$

- Duality: for the usual two-sided tests, a level α test does *not* reject $H_0 : \mu = \mu_0$ exactly when a $100(1-\alpha)\%$ CI for μ contains μ_0. (There is a similar relationship between one-sided tests and upper/lower confidence bounds.)

Particular tests: one population mean

- A test of $H_0 : \mu = \mu_0$ vs $H_A : \mu \neq \mu_0$ rejects when $|T| > C^*$, where

$$T = \frac{\bar{X} - \mu_0}{\sigma^*/\sqrt{n}} \; .$$

Here σ^* and C^* are as in the above discussion of two-sided confidence intervals.

- The p-value corresponding to $T = t$ can be found by looking up $P(|T| > t)$ for the normal distribution (when n is large) or the t_{n-1} distribution (when n is small). NOTE: because of the absolute value signs, you must multiply the tabled value by 2.

- The sample size n at which a two-sided level α test has power $1 - \beta$ under the alternative μ' is approximately

$$n = \left(\frac{\sigma^*(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'} \right)^2 \; .$$

In the one-sided case, put z_α in place of $z_{\alpha/2}$. (The resulting n is only valid if it is large, since the formula uses large-sample normality).

- A test of $H_0 : \mu \geq \mu_0$ vs $H_A : \mu < \mu_0$ would reject when $T < C^*$, where C^* is from the lower confidence bound case discussed above. p-values can be found analogously (do not multiply the tabled value by 2).

Particular tests: one population proportion

- To test $H_0 : p = p_0$ vs $H_A : p \neq p_0$, use

$$T = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \; ,$$

with the critical value determined in the usual manner as a standard normal upper quantile. p-values are also determined from $T = t$ in an analogous manner. Here, n should not be too small; $np_0(1 - p_0) > 5$ should suffice. For smaller n, there is a procedure we have not covered based on the binomial distribution.

- The n for which a two-sided test of p_0 has power $1 - \beta$ under the alternative $p = p'$ is approximately
$$n = \left(\frac{z_{\alpha/2}\sqrt{p_0(1 - p_0)} + z_\beta \sqrt{p'(1 - p')}}{p' - p_0} \right)^2 ,$$
rounded up to the nearest integer. For a one-sided test, replace $z_{\alpha/2}$ with z_α.

Chapter 9: Inferences based on two samples

Inferences about the difference of two population means

- A two-sided hypothesis test has the form $H_0 : \mu_1 - \mu_2 = \Delta_0$ vs $H_A : \mu_1 - \mu_2 \neq \Delta_0$. Often, $\Delta_0 = 0$. If the sample from population A is independent of the sample from population B, reject when $|T| > C^*$, where
$$T = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{(\sigma_1^*)^2/n_1 + (\sigma_2^*)^2/n_2}} .$$
Here σ_1^*, σ_2^*, and C^* are like the values in the one-sample procedures. However, if n_1 or n_2 is small and σ is unknown, you need to assume normal population distributions and treat T as having a t distribution with ν df. Here
$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}} ,$$
rounded to the nearest integer. Note that $\min\{n_1 - 1, n_2 - 1\} \le \nu \le n_1 + n_2 + 1$.

- A $100(1 - \alpha)\%$ CI for $\mu_1 - \mu_2$ takes the form $\bar{X} - \bar{Y} \pm C^* \cdot \mathrm{SE}$, where SE is the denominator of the test statistic T.

- p-values can be found in the usual way from the value $T = t$ and the corresponding table.

- Formulas for β can be derived from the structure of the test. Samples size calculations are complicated, except in special cases. We do not give general formulas here, or for two proportions below.

- If the additional assumption $\sigma_1 = \sigma_2$ is tenable, use
$$T = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{S^2_{\text{pooled}}(1/n_1 + 1/n_2)}} ,$$

where

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

- When each observation from sample A is paired with an observation from sample B, inferences are based on the differences $D_i = X_i - Y_i$. The two-sample test of $H_0 : \mu_1 - \mu_2 = \Delta_0$ then reduces to a one-sample test of $H_0 : \mu_D = \Delta_0$, with the D_i's as the sample.

Inferences about the difference of two population proportions

- When testing $H_0 : p_1 - p_2 = \Delta_0 \neq 0$, the situation is identical to the two-means case, except $(\hat{p}_1, \hat{p}_2)$ replaces $(\bar{X}, \bar{Y})$ and $\hat{p}_i(1 - \hat{p}_i)$ replaces $(\sigma_i^*)^2$, $i = 1, 2$.
- When testing $H_0 : p_1 - p_2 = 0$, use

  $$T = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)}} \ ,$$

  where $\hat{p}$ is the combined sample estimate of p, defined by

  $$\hat{p} = \frac{X + Y}{n_1 + n_2} = \frac{n_1}{n_1 + n_2} \hat{p}_1 + \frac{n_2}{n_1 + n_2} \hat{p}_2 \ .$$

  Here again, n_1 and n_2 should not be small.
- NOTE: To build a confidence interval for $p_1 - p_2$, you should not use the pooled variance estimate based on $\hat{p}$.

Chapters 10 and 11: ANOVA

The one-way layout

- The ith group mean is $\bar{Y}_{i\cdot} = (1/J_i) \sum_{j=1}^{J_i} Y_{ij}$, with J_i the size of the sample for group i.
- The grand mean is $\bar{Y}_{\cdot\cdot} = (1/n) \sum_{i=1}^{I} \sum_{j=1}^{J_i} Y_{ij}$, where $n = \sum_{i=1}^{I} J_i$.
- The total sum of squares is $\text{SST} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2$.
- The treatment (or model) sum of squares is $\text{SSTr} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$.
- The error (or residual) sum of squares is $\text{SSE} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{i\cdot})^2$.
- The sum of squares identity: $\text{SST} = \text{SSTr} + \text{SSE}$.
- The F statistic is $F = \text{MSTr}/\text{MSE} = [\text{SSTr}/(I - 1)]/[\text{SSE}/(n - I)]$. Its level-$\alpha$ critical value is $F_{\alpha; I-1; n-I}$.

The two-way layout

- Factor A has I levels, factor B has J levels, and we get K observations in each cell (i, j).
- The ith group mean for factor A is $\bar{Y}_{i\cdot\cdot} = 1/(JK)\sum_{j=1}^{J}\sum_{k=1}^{K} Y_{ijk}$.
- The jth group mean for factor B is $\bar{Y}_{\cdot j\cdot} = 1/(IK)\sum_{i=1}^{I}\sum_{k=1}^{K} Y_{ijk}$.
- The grand mean is $\bar{Y}_{\cdots} = 1/(IJK)\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} Y_{ijk}$.
- The total sum of squares is $\text{SST} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{\cdots})^2$.
- In the interaction model, the sums of squares are

$$\begin{aligned}
\text{SSA} &= \sum_i \sum_j \sum_k (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots})^2 \qquad \text{df } = I - 1 \\
\text{SSB} &= \sum_i \sum_j \sum_k (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots})^2 \qquad \text{df } = J - 1 \\
\text{SSAB} &= \sum_i \sum_j \sum_k (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})^2 \qquad \text{df } = (I-1)(J-1) \\
\text{SSE} &= \sum_i \sum_j \sum_k (\bar{Y}_{ijk} - \bar{Y}_{ij\cdot})^2 \qquad \text{df } = IJ(K-1)
\end{aligned}$$

- The test statistic for H_{0A} is $F_A = \text{MSA}/\text{MSE}$, with critical value $F_{\alpha;I-1;IJ(K-1)}$.
- The test statistic for H_{0B} is $F_B = \text{MSB}/\text{MSE}$, with critical value $F_{\alpha;J-1;IJ(K-1)}$.
- The test statistic for H_{0AB} is $F_{AB} = \text{MSAB}/\text{MSE}$, with critical value $F_{\alpha;(I-1)(J-1);IJ(K-1)}$.

Chapter 12: Simple linear least-squares regression

- The simple linear least-squares coefficient estimates are

$$\begin{aligned}
\hat{\beta}_1^{\text{LS}} &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\widehat{\text{Cov}}(X, Y)}{\widehat{\text{Var}}(X)} \\
\hat{\beta}_0^{\text{LS}} &= \bar{y} - \hat{\beta}_1 \bar{x}
\end{aligned}$$

- The estimate of σ^2 is $\hat{\sigma}^2 = S^2 = [\sum_i (y_i - \hat{y})^2]/(n-2)$.
- The coefficient of determination is $R^2 = 1 - \text{SSR}/\text{SST}$, where SSR is the same residual sum of squares used to estimate σ^2, and SST is the total sum of squares $\sum_i (y_i - \bar{y})^2$.

- Let $s_{xx} = \sum_i (x_i - \bar{x})^2$. Then $\mathrm{Var}(\hat{\beta}_1) = \sigma^2/s_{xx}$. The estimated SE is $S_{\hat{\beta}_1} = S/\sqrt{s_{xx}}$.

- Under the ideal linear model, $(\hat{\beta}_1 - \beta_1)/S_{\hat{\beta}_1}$ has the t_{n-2} distribution. This gives a test and a CI in the usual way.

- $\mathrm{Var}(\hat{\beta}_0) = \sigma^2 \frac{(\sum x_i^2)/n}{s_{xx}}$. We get another estimated SE, another t_{n-2} pivot, and another test/CI.

- Let $\hat{Y}$ be the estimated mean $\hat{\beta}_0 + \hat{\beta}_1 x^*$ at a fixed point x^*. The estimated SE of $\hat{Y}$ is

  $$S_{\hat{Y}} = S\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}}\ ,$$

  and we have

  $$\frac{\hat{Y} - (\beta_0 + \beta_1 x^*)}{S_{\hat{Y}}} \sim t_{n-2}\ .$$

  This leads to a test/CI in the usual way. Using $\sqrt{S^2 + S_{\hat{Y}}^2}$ instead of $S_{\hat{Y}}$ gives a PI instead of a CI.

Chapter 13: Nonlinear and multiple regression

- In multiple regression, to test the null hypothesis that all the partial slope coefficients are zero, we have the test statistic $F = (\mathrm{SSM}/k)/(\mathrm{SSR}/(n-k-1))$. SSM is the model sum of squares, i.e., $\mathrm{SST} - \mathrm{SSR}$. SSR is the residual sum of squares. k is the number of partial slope coefficients (the number of predictor variables). The critical value is $F_{\alpha;k;n-k-1}$.

- To test the null hypothesis that $\beta_{l+1}, \ldots, \beta_k$ are all zero, the test statistic is $F = [(\mathrm{SSR}_l - \mathrm{SSR}_k)/(k-l)]/[\mathrm{SSR}_k/(n-k-1)]$. SSR_l and SSR_k are, respectively, the residual sums of squares for the model fit to the first l predictors, and to all k predictors. The critical value is $F_{\alpha;k-l;n-k-1}$.

Chapter 14: Categorical data analysis and goodness of fit

- The χ^2 statistic to test goodness-of-fit of outcome counts $N_1, \ldots, N_k$ to specified outcome probabilities $p_1, \ldots, p_k$ is

  $$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - E_i)^2}{E_i}\ ,$$

  where $E_i = np_i$ and n is the number of trials for the multinomial. The critical value is $\chi^2_{\alpha;k-1}$.

- If the outcome probabilities are functions of a parameter θ, so that $p_1 = \pi_1(\theta)$, etc., then $H_0 : \theta = \theta_0$ is tested in the same way, using outcome probabilities $\pi_1(\theta_0), \ldots, \pi_k(\theta_0)$. Here θ_0 is a distinguished value that we know.

- If the outcome probabilities are parametric functions, but we only know θ lies in a set Θ, without knowing its specific value, then

  - we compute a maximum-likelihood estimate $\hat{\theta}$ based on the counts
  - we do the χ^2 test using outcome probabilities $\pi_1(\hat{\theta}), \ldots, \pi_k(\hat{\theta})$.

  The critical value of the test is $\chi^2_{\alpha;k-1-m}$, where m is the number of components in θ. If θ is a scalar, $m = 1$.